Lecture #7: Regular Operations and Regular Expressions
Key Concepts

Regular Operations

This lecture introduced the ***regular operations*** (union, concatenation, and Kleene star) on languages. Let Σ be an alphabet and let $A, B \subseteq \Sigma^\star$, so that A and B are *languages* with alphabet Σ.

- The ***union*** of the languages A and B is the language

$$A \cup B = \{\omega \in \Sigma^\star \mid \omega \in A \text{ or } \omega \in B \text{ (or both)}\}.$$

- The ***concatenation*** of the languages A and B is the language

$$A \circ B = \{\omega_1 \cdot \omega_2 \mid \omega_1 \in A \text{ and } \omega_2 \in B\}.$$

- The ***Kleene star*** of the language A is the language

$$A^\star = \{\omega_1 \cdot \omega_2 \ldots \omega_k \mid k \geq 0 \text{ and } \omega_i \in A \text{ for } 1 \leq i \leq k\}$$

  This language is also, sometimes called the ***Kleene closure*** of A — or the ***star*** of A.

Closure Properties

A ***closure property*** for a set S of languages over an alphabet Σ, is a property stating — for an *operation* on languages over Σ — that if the operation is applied to languages that all belong to the set S, then the result is a language that belongs to the set S, as well. The following result describes several closure properties for the set of regular languages.

Theorem 1. *Let Σ be an alphabet, and let $A, B \subseteq \Sigma^\star$.*

(a) If A and B are regular languages then $A \cup B$ is a regular language, as well.

(b) If A and B are regular languages, then $A \circ B$ is a regular language, as well.

(c) If A is a regular language then $A^\star$ is a regular language as well.

Regular Expressions and Their Languages

Let Σ be an alphabet that *does not* include any of the symbols

$$\lambda, \emptyset, \Sigma, (,), \cup, \circ, ^\star$$

and let

$$\Sigma_{\text{regexp}} = \Sigma \cup \{\lambda, \emptyset, \text{“}\Sigma\text{”}, (,), \cup, \circ, ^\star\}$$

so that Σ_{regexp} incudes a copy of the *symbol*, “Σ”, that we are also using as the name of the language we are starting with.

A ***regular expression over the alphabet*** $\boldsymbol{\Sigma}$ is a kind of string of symbols, in $\Sigma_{\text{regexp}}^\star$, as defined by the following list of seven rules — and he ***language of a regular expression, over the alphabet*** $\boldsymbol{\Sigma}$, is a subset of $\Sigma^\star$ that is defined as follows, as well.

1. For every ***symbol*** $\sigma \in \Sigma$, the ***string*** σ, with length one in $\Sigma_{\text{regexp}}^\star$ is a regular expression over Σ. The ***language***, $L(\sigma)$, of the regular expression σ, is the ***set*** $\{\sigma\}$.

2. The ***string*** λ, with length one in $\Sigma_{\text{regexp}}^\star$, is a regular expression over Σ. The ***language***, $L(\lambda)$, of the regular expression λ, is the ***set*** $\{\lambda\}$.

3. The ***string*** $\emptyset$, with length one in $\Sigma_{\text{regexp}}^\star$ is a regular expression over Σ. The ***language***, $L(\emptyset)$, of the regular expression $\emptyset$, is the ***set*** $\emptyset$.

4. The ***string*** Σ, with length one in $\Sigma_{\text{regexp}}^\star$, is a regular expression over (the alphabet) Σ. The ***language***, $L(\Sigma)$, of the regular expression Σ, is the ***finite set*** Σ.

5. If $R_1 \in \Sigma_{\text{regexp}}^\star$ and $R_2 \in \Sigma_{\text{regexp}}^\star$ are ***regular expressions over*** $\boldsymbol{\Sigma}$ then the ***string***

   $$(\, R_1 \,\cup\, R_2 \,) \tag{1}$$

   (of symbols in Σ_{regexp}) is a regular expression over Σ. The ***language*** $L(R)$, of the regular expression R at line (1), is the ***set***

   $$L(R_1) \cup L(R_2).$$

6. If $R_1 \in \Sigma_{\text{regexp}}^\star$ and $R_2 \in \Sigma_{\text{regexp}}^\star$ are ***regular expressions over*** $\boldsymbol{\Sigma}$ then the ***string***

$$(R_1 \circ R_2) \tag{2}$$

(of symbols in Σ_{regexp}) is a regular expression over Σ. The ***language*** $L(R)$, of the regular expression R at line (2), is the ***set***

$$L(R_1) \circ L(R_2).$$

7. If $R_1 \in \Sigma_{\text{regexp}}^\star$ is a ***regular expression over*** $\boldsymbol{\Sigma}$ then the ***string***

$$(R_1)^\star \tag{3}$$

(of symbols in Σ_{regexp}) is a regular expression over Σ. The ***language*** $L(R)$, of the regular expression R at line (3), is the ***set***

$$L(R_1)^\star$$

Theorem 2. *Let Σ be an alphabet that does not include any of the symbols "λ", "$\emptyset$", "Σ", "$($", "$)$", "$\cup$", "$\circ$", or "$\star$", and let $L \subseteq \Sigma^\star$.*

Then L is a **regular language** *if and only if there exists a regular expression R, over the alphabet Σ, such that $L = L(R)$.*

Note: Regular expressions are strings of text that can be used to describe languages in pseudocode and code — and these have significant applications in system software and knowledge representation. If there was more time in this course, then quite a bit of the extra time would be devoted to regular expressions and their processing. Some of the information that would be included, if time was available, is given in the *optional* supplemental documents for this lecture.