Regular Languages (Section 11.1)
Regular Languages

• **Problem**: Suppose the input strings to a program must be strings over the alphabet \{a, b\} that contain exactly one substring \(bb\). In other words, the strings must be of the form \(xbby\), where \(x\) and \(y\) are strings over \{a,b\} that do not contain \(bb\), \(x\) does not end in \(b\), and \(y\) does not begin with \(b\). Later we will see how to describe such strings formally.

• A **regular language** over the alphabet \(A\) is a language constructed by the following rules:
  
  - \(\emptyset\) and \(\{\Lambda\}\) are regular languages.
  - \(\{a\}\) is a regular language for all \(a \in A\).
  - If \(M\) and \(N\) are regular languages, then so are \(M \cup N\), \(MN\), and \(M^*\).
Example: Let $A = \{a, b\}$. Then the following languages are a sampling of the regular languages over $A$:

- $\emptyset$, {$\Lambda$}, {$a$}, {$b$}, {$a, b$}, {$ab$}, {$a$}*, {$\Lambda, a, aa, aaa, \ldots, a^n, \ldots$}
• A regular expression over alphabet $A$ is an expression constructed by the following rules:
  • $\emptyset$ and $\Lambda$ are regular expressions.
  • $a$ is a regular expression for all $a \in A$.
  • If $R$ and $S$ are regular expressions, then so are $(R), R + S, R \cdot S,$ and $R^*.$

The hierarchy in the absence of parentheses is: $\ast$ (do it first), $\cdot,$ $+$ (do it last). Juxtaposition will be used instead of $\cdot.$

• Example: Let $A = \{a, b\}.$ Then the following expressions are a sampling of the regular expressions over $A$:
  • $\emptyset, \Lambda, a, b, ab, a + ab, (a + b)^*$
Regular expressions represent regular languages by the following correspondence, where $L(R)$ denotes the regular language of the expression $R$:

- $L(\emptyset) = \emptyset$
- $L(\Lambda) = \{\Lambda\}$
- $L(a) = \{a\}$ for all $a \in A$
- $L(R + S) = L(R) \cup L(S)$
- $L(RS) = L(R)L(S)$
- $L(R^*) = L(R)^*$
Example: The regular expression $ab + a^*$ represents the following regular language:

$L(ab + a^*) = L(ab) \cup L(a^*)$

$= L(a)L(b) \cup L(a)^*$

$= \{a\}\{b\} \cup \{a\}^*$

$= \{ab\} \cup \{\Lambda, a, aa, aaa, \ldots, a^n, \ldots\}$

$= \{ab, \Lambda, a, aa, aaa, \ldots, a^n, \ldots\}$
Example: The regular expression \((a + b)^*\) represents the following regular language:

- \(L((a + b)^*) = (L(a + b))^* = \{a, b\}^*\), the set of all possible strings over \(\{a, b\}\).
• Back to the problem: Suppose the input strings to a program must be strings over the alphabet \{a, b\} that contain exactly one substring \textit{bb}. In other words, the strings must be of the form \textit{xbby}, where \textit{x} and \textit{y} are strings over \{a, b\} that do not contain \textit{bb}, \textit{x} does not end in \textit{b}, and \textit{y} does not begin with \textit{b}. How can we describe the set of inputs formally?

Solution: Let \textit{x} = \(a + ba\)^* and \textit{y} = \(a + ab\)^*. So the entire answer is: \((a + ba)^* bb (a + ab)^*\)
• Back to the problem: Suppose the input strings to a program must be strings over the alphabet \{a, b\} that contain exactly one substring \textit{bb}. In other words, the strings must be of the form \textit{xbby}, where \textit{x} and \textit{y} are strings over \{a, b\} that do not contain \textit{bb}, \textit{x} does not end in \textit{b}, and \textit{y} does not begin with \textit{b}. How can we describe the set of inputs formally?

• Solution: Let \textit{x} = (a + ba)^* \textit{ and } \textit{y} = (a + ab)^*.
• *Back to the problem:* Suppose the input strings to a program must be strings over the alphabet \{a, b\} that contain exactly one substring \(bb\). In other words, the strings must be of the form \(xbby\), where \(x\) and \(y\) are strings over \{a, b\} that do not contain \(bb\), \(x\) does not end in \(b\), and \(y\) does not begin with \(b\). How can we describe the set of inputs formally?

• *Solution:* Let \(x = (a + ba)^*\) and \(y = (a + ab)^*\).

• So the entire answer is: \((a + ba)^* bb(a + ab)^*\)
The Algebra of Regular Expressions

- **Equality:** Regular expressions $R$ and $S$ are equal, written $R = S$, when $L(R) = L(S)$.
- **Examples:** $a + b = b + a$, $a + a = a$, $aa^* = a^*a$, $ab \neq ba$
- **Properties of $+$, $\cdot$, and closure:**
  - $+$ is commutative, associative, $\emptyset$ is identity for $+$, and $R + R = R$.
  - $\cdot$ is associative, $\Lambda$ is identity for $\cdot$, and $\emptyset$ is zero for $\cdot$.
  - $\cdot$ distributes over $+$
Simplify the regular expression $aa(b^* + a) + a(ab^* + aa)$

$aa(b^* + a) + a(ab^* + aa)$

$= aa(b^* + a) + aa(b^* + a) \quad \cdot \text{distributes over } +$

$= aa(b^* + a)$

$R + R = R.$
Show that \((a + aa)(a + b)^* = a(a + b)^*\)

\[
\begin{align*}
(a + aa)(a + b)^* & = (a + aa)a^*(ba^*)^* \\
& = a(\Lambda + a)a^*(ba^*)^* \\
& = aa^*(ba^*)^* \\
& = a(a + b)^* \\
QED
\end{align*}
\]