Regular expressions I

- Regular expressions are a convenient means of specifying certain simple sets of strings.
- Each regular expression \( r \) denotes a language \( L(r) \)
- A language denoted by a regular expression is said to be a regular set.
- The defining rules specify how \( L(r) \) is formed by combining in various ways the language denoted by the subexpressions of \( r \).
Regular expressions II

The rules that define the regular expressions over alphabet $\Sigma$ are following:

1. $\varepsilon$ is a regular expression that denotes $\{\varepsilon\}$, that is, the set containing the empty string.

2. If $a$ is a symbol in $\Sigma$, then $a$ is a regular expression that denotes $\{a\}$, i.e., the set containing the string $a$. Although we use the same notation for all three, technically, the regular expression $a$ is different from the string $a$ or the symbol $a$. It will be clear from the context whether we are talking about $a$ as a regular expression, string, or symbol.

3. Suppose $r$ and $s$ are regular expressions denoting the language $L(r)$ and $L(s)$. Then
   - $(r)|(s)$ is a regular expression denoting $L(r) \cup L(s)$
   - $(r)(s)$ is a regular expression denoting $L(r)L(s)$
   - $(r)^*$ is a regular expression denoting $(L(r))^*$
   - $(r)^2$ is a regular expression denoting $L(r)^2$
Let $\Sigma = (a,b)$

1. The regular expression $a|b$ denotes the set $\{a,b\}$.
2. The regular expression $(a|b)(a|b)$ denotes $\{aa, ab, ba, bb\}$, the set of all strings of $a$’s and $b$’s of length two.
3. The regular expression $a^*$ denotes the set of all strings containing zero or more $a$’s, i.e., $\{\epsilon, a, aa, aaa, \ldots\}$
4. The regular expression $(a|b)^*$ denotes the set of all strings containing zero or more instances of an $a$ or $b$, that is, the set of all strings of $a$’s and $b$’s.
5. The regular expression $a|a^*b$ denotes the set containing the string $a$ and all strings containing of zero or more $a$’s followed by a $b$. 