Regular Expression

- a *regular expression* is a specific kind of pattern that describes strings with a certain form

Regular Expressions

Definition 3.2. Let Σ be an alphabet. Then the following patterns are ***regular expressions*** over Σ:

1. Φ and ε are regular expressions;
2. a is a regular expression, for each $a \in \Sigma$;
3. if r_1 and r_2 are regular expressions, then so are $r_1 \,|\, r_2$, $r_1 \cdot r_2$, r_1^* and (r_1) (and of course, r_2^* and (r_2)). As in concatenation of strings, the $\cdot$ is often left out of the second expression. (Note: the order of precedence of operators, from lowest to highest, is $|$, $\cdot$, $*$.)

No other patterns are regular expressions.

fee or fi? the Greek pronunciation of Φ is fee, but fi is common in (US) English (and math)

- so far this only describes the syntax of a regular expression – what sequences of symbols one can write down to form a regular expression

Regular Expressions

Definition 3.3. The ***language generated by a regular expression*** r, denoted $L(r)$, is defined as follows:

1. $L(\Phi) = \emptyset$, i.e. no strings match Φ;
2. $L(\varepsilon) = \{\varepsilon\}$, i.e. ε matches only the empty string;
3. $L(a) = \{a\}$, i.e. a matches only the string a;
4. $L(r_1 \,|\, r_2) = L(r_1) \cup L(r_2)$, i.e. $r_1 \,|\, r_2$ matches strings that match r_1 or r_2 or both;
5. $L(r_1 r_2) = L(r_1)L(r_2)$, i.e. $r_1 r_2$ matches strings of the form "something that matches r_1 followed by something that matches r_2";
6. $L(r_1^*) = (L(r_1))^*$, i.e. r_1^* matches sequences of 0 or more strings each of which matches r_1.
7. $L((r_1)) = L(r_1)$, i.e. (r_1) matches exactly those strings matched by r_1.

- this defines what a given regular expression means

Regular Languages

- a language is *regular* if it is generated by a regular expression

- the union of two regular languages is regular ✔
- the concatenation of two regular languages is regular ✔
- the Kleene closure of a regular language is regular ✔

- the intersection of two regular languages is regular ?
- the complement of a regular languages is regular ?

1. Give English-language descriptions of the languages generated by the following regular expressions.

a) $(a \mid b)^*$ **b)** $a^* \mid b^*$ **c)** $b^*(ab^*ab^*)^*$ **d)** $b^*(abb^*)$

2. Give regular expressions over $\Sigma = \{a, b\}$ that generate the following languages.

a) $L_1 = \{x \mid x \text{ contains 3 consecutive } a\text{'s}\}$
b) $L_2 = \{x \mid x \text{ has even length}\}$
c) $L_3 = \{x \mid n_b(x) = 2 \bmod 3\}$
d) $L_4 = \{x \mid x \text{ contains the substring } aaba\}$
e) $L_5 = \{x \mid n_b(x) < 2\}$
f) $L_6 = \{x \mid x \text{ doesn't end in } aa\}$

3. Prove that all finite languages are regular.