

Regular expressions

ian.mcloughlin@gmit.ie

Regular expressions

$(0.0|1.1).(0.1)^*.(0.0|1.1)$

- Regular expressions are strings that represent patterns of text.
- The strings can contain special characters.
- Brackets can be used to group characters together.
- Regular expressions are used to search other strings for patterns.

Special characters

- $.$ means *concatenate*. So, $a.b$ means an a followed by a b .
- $|$ means *or*. So, $a|b$ means an a or a b .
- $*$ means *zero or more times*. So, a^* means zero or more a 's.

Examples of regular expressions

$a.b.c$ – a single a followed by a single b followed by a single c .

$a.b.c^*$ – an a followed by a b followed by zero or more c 's.

$a|b.c$ – an a , or a b followed by a c .

$(a|b).c$ – an a or a b , followed by a c .

$0.0.(0|1)^*$ – all strings of 0's and 1's that begin with two zeros.

1^* – any number of 1's (including empty string).

Precedence

1. Always apply $*$ first.
2. Apply $.$ after $*$ but before $|$.
3. Apply $|$ last.
4. Treat bracketed groups as individual characters.

Languages of regular expressions

This is a list of regular expressions over an alphabet A and the languages they define.

\emptyset – the empty language ($\{\}$).

ϵ – the language containing the empty string ($\{\epsilon\}$).

a – the language containing any symbol a from A ($\{a\}$).

$r.s$ – the concatenation of the languages defined by regular expressions r and s .

$r|s$ – the union of the languages of regular expressions r and s .

r^* – the star of the language defined by regular expression r .

Infix and postfix

It is sometimes convenient to re-write expressions in postfix. This applies to lots of different expressions, not just regular expressions.

Example

Convert an infix mathematical expression (left) to postfix (right):

$$(3 + 4) \times 5 \quad \rightarrow \quad 3\ 4\ +\ 5\ \times$$

Example

Converting an infix regular expression (left) to postfix (right):

$$a.(b.b)^*.a \quad \rightarrow \quad abb.*.a.$$

Note we often omit the `.` in infix notation: “`a(bb)*a`” but can’t in postfix. However, the brackets aren’t needed in postfix.