

Regular Expressions

David Robinson

Regular Expressions

Components of a regular expression describing a language over the alphabet Σ :

- A symbol to represent the language containing the string consisting of itself
- $a \cup b$ to represent either of symbols a or b
- $a \circ b$ or just ab to represent symbol a concatenated with symbol b
- Σ to represent any symbol from Σ
- a^* to represent zero or more occurrences of a
- Σ^* to represent zero or more occurrences of any symbol from Σ

R is a **regular expression** over the alphabet Σ if it is:

- a for some $a \in \Sigma$
- ε
- θ
- $R_1 \cup R_2$ where R_1 and R_2 are both regular expressions
- $R_1 \circ R_2$ where R_1 and R_2 are both regular expressions
- R_1^* where R_1 is a regular expression

Generalized Nondeterministic Finite Automaton

A generalized nondeterministic finite automaton (GNFA) is a special kind of NFA that uses regular expressions as its transition alphabet. It has a single start state and a single accept state.

A GNFA is a 5-tuple $G = \{Q, \Sigma, \delta, q_s, q_f\}$ where:

- Q is the set of states,
- Σ is the input alphabet
- $\Delta : (Q - \{q_s\}) \times (Q - \{q_f\}) \rightarrow \mathbf{R}$ is the transition function with \mathbf{R} being the set of all regular expressions over Σ
- q_s is the start state
- q_f is the accept state