Graph unit

- Languages, finite automata, and regular expressions (**today**)
- Equivalence of models of computation (Friday)
- The lambda calculus (next week Monday)
- Computability and tractability (next week Wednesday)

Today:

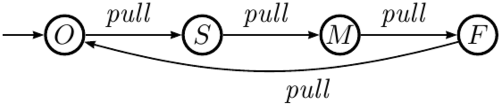
- Languages
- Finite automata
- Nondeterminism
- Regular expressions

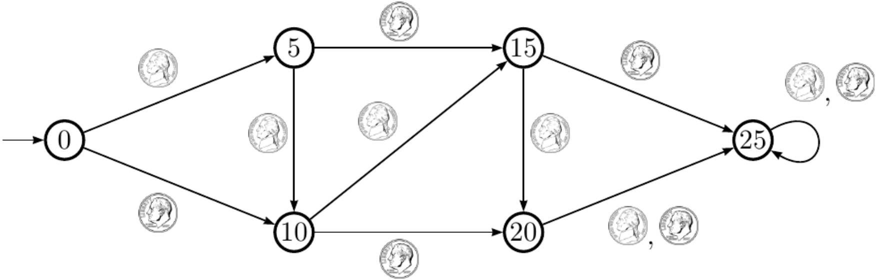
alphabet	$\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$	$\{A, B, \Gamma, \Delta\}$	{ ☎, ⬢, ☕, ⌧ }
some strings	ε	$AA\Gamma\Delta BAA$	☎☎
over the	aaaabaaabcaaaa	$\Gamma\Gamma\Gamma$	☎⌧⬢⌧
alphabet	bbabc	$\Delta\Gamma AA\Delta$	☕☕☎☕⬢☕
	ccccc	ε	⬢⬢⌧☕⌧⬢⬢

A language over Σ is any subset of $\Sigma*$. Let $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, L_1 be the language of strings of size 3, L_2 the language of strings with only occurrences of a, and L_3 be the language of palindromes (strings the same backwards and forwards), all over the alphabet Σ.

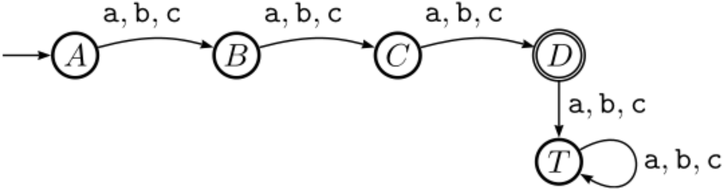
$$\begin{array}{rcl} \mathtt{aaa} & \in & L_1 \\ \mathtt{aba} & \in & L_1 \\ \mathtt{abc} & \in & L_1 \\ \mathtt{ccb} & \in & L_1 \\ \mathtt{cbb} & \in & L_1 \end{array} \qquad \begin{array}{rcl} \mathtt{aaa} & \in & L_2 \\ \mathtt{a} & \in & L_2 \\ \mathtt{aaaaaaaaa} & \in & L_2 \\ \mathtt{aa} & \in & L_2 \\ \varepsilon & \in & L_2 \end{array} \qquad \begin{array}{rcl} \mathtt{aaa} & \in & L_3 \\ \mathtt{aba} & \in & L_3 \\ \mathtt{aabaa} & \in & L_3 \\ \mathtt{aaccaa} & \in & L_3 \\ \varepsilon & \in & L_3 \end{array}$$

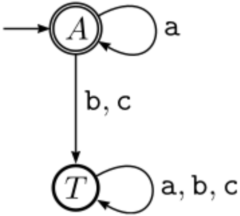
$$\begin{array}{rcl} L_1 & = & \{x \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}* \mid |x| = 3\} \\ \\ L_2 & = & \{\mathtt{a}^n \mid n \geq 0\} \text{ or simply } \{a\}* \\ \\ L_3 & = & \{xyx^R \mid x, y \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}*,\ |y| = 0 \text{ or } |y| = 1\} \end{array}$$

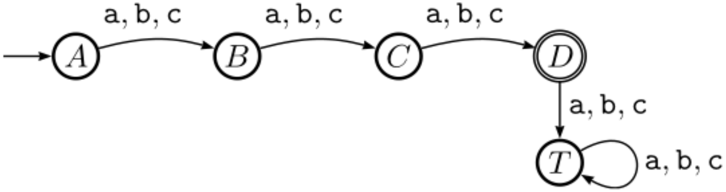




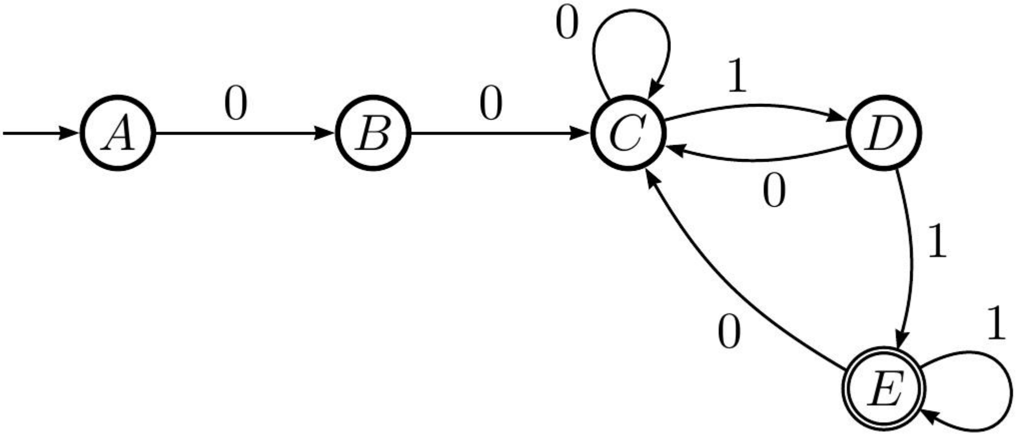
Diagrams generated by VAUCANSON-G


A
B
C
D
T
a, b, c
a, b, c
a, b, c
a, b, c
a, b, c

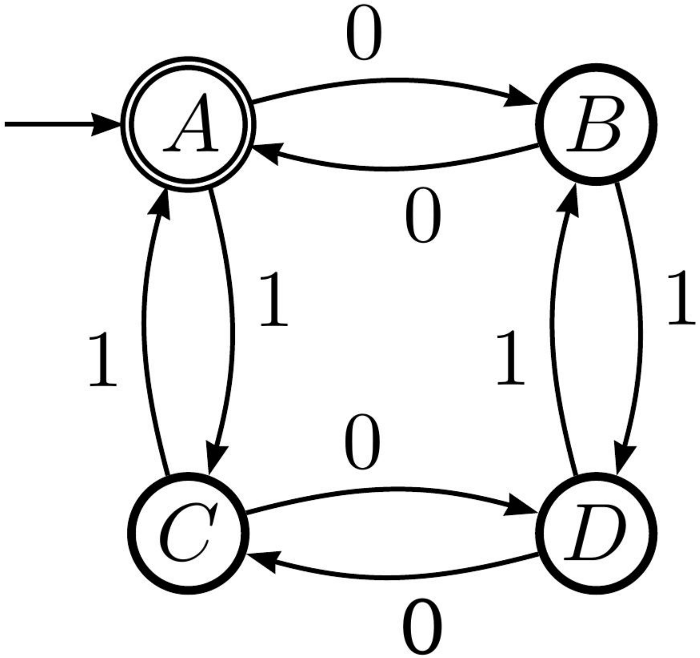

A
T
a
b, c
a, b, c



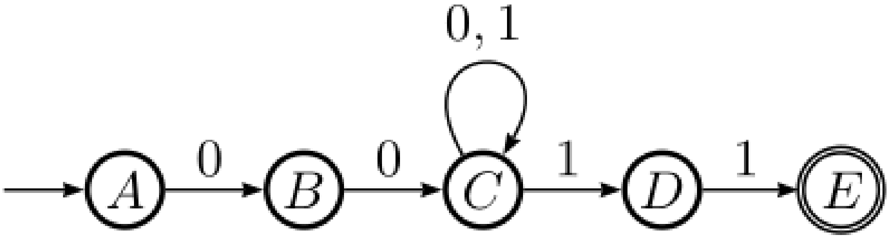
$$
\begin{aligned}
Q &= \{A, B, C, D, T\} \\
\Sigma &= \{\mathtt{a}, \mathtt{b}, \mathtt{c}\} \\
\delta &= \{((\mathtt{a}, A), B), (\mathtt{b}, A), B), (\mathtt{c}, A), B), \\
&\quad ((\mathtt{a}, B), C), (\mathtt{b}, B), C), (\mathtt{c}, B), C), \\
&\quad ((\mathtt{a}, C), D), (\mathtt{b}, C), D), (\mathtt{c}, C), D), \\
&\quad ((\mathtt{a}, D), T), (\mathtt{b}, D), T), (\mathtt{c}, D), T), \\
&\quad ((\mathtt{a}, T), T), (\mathtt{b}, T), T), (\mathtt{c}, T), T)\} \\
q_0 &= A \\
F &= \{D\}
\end{aligned}
$$



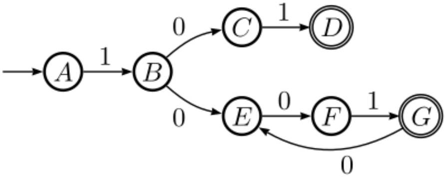
A
B
C
D
E
0
0
0
1
0
1
0
1



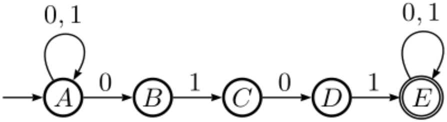
A
B
C
D
0
0
0
0
1
1
1
1


0, 1
0
0
1
1
A
B
C
D
E

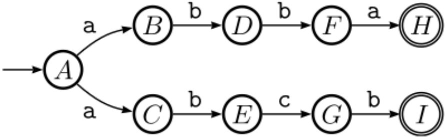
$\Sigma = \{0, 1\}$. Strings beginning with 1 followed by occurrences of 001, with the special case that string 101 is in the language. Examples: 101, 1001, 1001001.



$\Sigma = \{0, 1\}$. Strings containing the substring 0101. Examples: 01011111, 000101001, 0010111010100.



$\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. The language of just the strings `abba` and `abcb`.



Given an alphabet Σ, a **regular expression** is one of the following:

- $\emptyset$, which denotes the set $\emptyset$, that is, the set/language of no strings.
- ε, which denotes the set $\{\varepsilon\}$, that is, the set/language containing only the empty string.
- $\mathtt{a}$, where $\mathtt{a} \in \Sigma$, which denotes the set $\{\mathtt{a}\}$, that is, the set/language containing only the string with only the symbol $\mathtt{a}$.
- $r|s$, where r and s are regular expressions, which denotes the set/language $r \cup s$ (remember that r and s each denote a set). For example, $\mathtt{a}|\varepsilon$ represents the language $\{\mathtt{a}, \varepsilon\}$, and $\mathtt{a}|\mathtt{b}|\mathtt{c}$ represents the language $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$.
- rs, where r and s are regular expressions, which denotes the set/language of strings each composed of a string from r concatenated with a string from s; formally, $\{RS \mid R \in r \text{ and } S \in s\}$. For example, $(\mathtt{a}|\mathtt{b}|\varepsilon)\mathtt{c}(\mathtt{d}|\mathtt{e})$ represents the language $\{\mathtt{acd}, \mathtt{ace}, \mathtt{bcd}, \mathtt{bce}, \mathtt{cd}, \mathtt{ce}\}$.
- $r*$, where r is a regular expression, denoting the set of strings which are composed of the concatenation of zero or more strings, each from r; formally, $\{R_0 R_1 \ldots R_n \mid n \in \mathbb{W} \text{ and } \forall\ i, 0 \leq i \leq n, R_i \in r\}$. For example, $\mathtt{a}(\mathtt{bc}) * \mathtt{d}$ represents the language $\{\mathtt{ad}, \mathtt{abcd}, \mathtt{abcbcd}, \mathtt{abcbcbcd}, \ldots\}$.

For Fri, Apr 24:

Take quiz (that doesn't exist yet) on Canvas