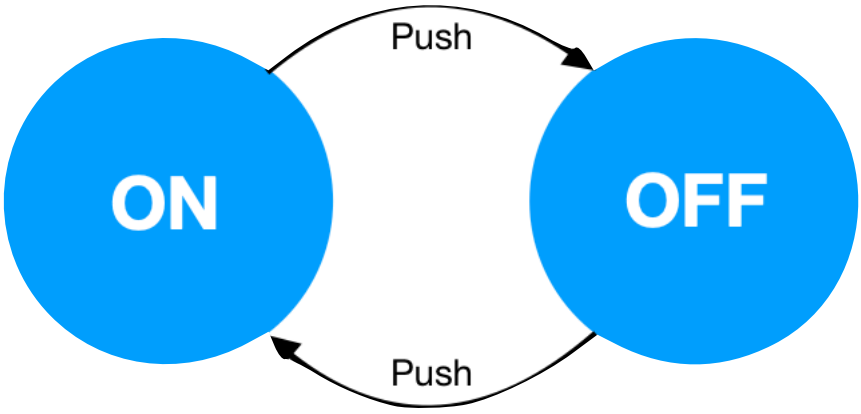
Automata and Formal Languages

Subject

- **What is a computer**, what can/can't be computed, what can be computed efficiently?
- RegEx, automata, grammars

Finite State Machine (Finite Automaton)

Machines with a finite amount of memory (finite amount of states)



Sets: $\{o_1, o_2, \ldots, o_n\}$, o_i is a member of the set

$S = \{1,2,3\} \implies 1 \in S, \qquad 4 \notin S$

Subset, when a set is all contained within a greater one $A \subseteq B$

$\{n \mid$ rule about n$\}$ to describe all kinds of n: $\{n \mid n \in \mathbb{Z} \land n$ is odd$\}$

Union $A \cup B$, **Intersection** $A \cap B$, **Complement** $\overline{A}$

Power set is a set of all possible sets (combination):

$A = \{0,1\} \implies$ power set of $A = \{\varnothing, \{0\}, \{1\}, \{0,1\}\}$

A **sequence** is a list of objects in some order, displayed with round parenthesis:

$(1,2,6)$

A **tuple** is a finite sequence

K-tuple is a sequence of k objects

The Cartesian product of A and B: $A \times B$

A set of all pairs where the first element is a member of A and the second is a member of B.

A **function** is a n object that sets an input-output relationship $f(a) = b$

All the possible inputs to the function is called the **domain**, whereas the outputs are the **range**:

Addition for integers is a function: $\mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{Z}$

An **alphabet** is a finite set of **symbols**

$\Sigma_1 = \{0,1\}; \Sigma_2 = \{a, b, c, d, e, \ldots, z\}$

A **string** over an alphabet is **finite sequence** of symbols from an alphabet

Hello $= \{H, e, l, l, o\}$ string over Σ_2

A **language** is a set of strings

$E = \{\text{hello, world, sun}\}$ is a language over Σ_2

Natural languages are the same as formal languages, sets of strings over an alphabet.

Powers of an alphabet: the set of all strings of a certain length over an alphabet (using exponential notation)

Σ^k is the set of strings of length k, each of whose symbols is in Σ

$\Sigma^0 = \{\epsilon\}$, ϵ is a string of length 0 by definition: $|\epsilon| = 0$ called the **empty string**.

The set of all strings over an alphabet Σ is denoted as Σ^*

$\Sigma^* = \Sigma^0 + \Sigma^1 + \Sigma^2 + \ldots$

For any finite Automaton:

5-tuple $(Q, \Sigma, \delta, q_0, F)$

Q is the set of **states**

Σ is an alphabet

$\delta : Q \times \Sigma \mapsto Q$ is the transition function

$q_0 \in Q$ is the start state and

$F \subseteq Q$ is the set of accept states

If A is the set of all strings that machine M accepts $\Longrightarrow$ A is the language of M

M recognises A