Università
Ca'Foscari
Venezia

Corso di Laurea magistrale (*ordinamento ex D.M. 270/2004*)
in Informatica - Computer Science

Tesi di laurea

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Duality and Regular Languages

Relatore
Ch. Prof. Antonino Salibra

Laureando
Riccardo Dalla Mora
Matricola 810336

Anno Accademico
2013/2014

dedicated to my family and my friends

Contents

Introduction **ix**

1 Topology and set theory **1**
1.1 Sets and relations 1
1.2 Topology 1
1.3 Metric spaces 3

2 Algebras **5**
2.1 Algebraic notions 5
2.2 Semigroups and monoids 8

3 Formal languages **11**
3.1 Formal languages theory 11
3.2 Chomsky classification 11
3.3 Regular languages 14
3.3.1 Language operations 14
3.3.2 Regular expressions 15
3.4 Finite automata and regular languages 16
3.5 Recognisable sets 18

4 Profinite topology and duality **21**
4.1 Profinite topology and profinite words 21
4.2 Connections to languages 25
4.3 Duality of Boolean algebras 26
4.4 Duality of regular languages 30

5 Conclusions **33**

References **41**

Abstract

In this master thesis we analyse the class of regular languages from an algebraic point of view. There exist interesting relations between regular languages and algebraic structures like monoids, Boolean algebras, lattices and varieties of algebras. In particular we focus on finite monoids. In fact, a language L is regular iff there exists a homomorphism from the free monoid $\mathbf{A}^*$ of the words onto a finite monoid in such a way that L is the inverse image of a subset of the monoid. Monoids can be used to define a metric over A^*, whose completion defines the Boolean space of profinite words, where a profinite word is a Cauchy sequence of words of A^*. The Priestly duality between the Boolean space of profinite words and the Boolean algebra of its clopen subsets can be used to prove the following theorems:

1. A class of languages is a Boolean algebra of languages iff the class can be defined by a set of profinite equations;
2. A language is regular iff its closure is clopen in the Boolean space of profinite words;
3. String concatenation is the dual operation of the residuals of right/left product on the Boolean algebra of regular languages.

Introduction

In this master thesis we focus on the main mathematical consequences which arise by analysing regular languages from an algebraic point of view. By Kleene Theorem the class of regular languages, the least class of languages cointaining the finite languages and closed under union, concatenation and Kleene star, coincides with the class of languages recognised by a finite automaton. Every finite automaton defines an observational equivalence over the free monoid $\mathbf{A}^*$ of the words over a finite alphabet A. Roughly speaking, two words are observationally equivalent iff there is no way to distinguish them by using the given automaton. The observational equivalence $\equiv$ is a congruence on $\mathbf{A}^*$ with a finite number of equivalence classes (so that the quotient $A^*/\equiv$ is a finite monoid) and the language recognised by the automaton is union of some of these equivalence classes. Finite monoids can be directly used to recognise languages: a language L is regular iff there exists a homomorphism from $\mathbf{A}^*$ onto a finite monoid such that L is the inverse image of a subset of the monoid. The syntactic monoid of a regular language L is the monoid of minimal cardinality recognising L.
Interesting results can be obtained by topologically characterising languages. An ultrametric d on A^* can be defined as follows, for all $x, y \in A^*$: $d(x,y) = 2^{-r(x,y)}$, where $r(x,y)$ is the cardinality of the least finite monoid separating x and y. Although the topology generated by (A^*, d) is the discrete topology, the completion $(\widehat{A^*}, d^*)$ of the metric space (A^*, d) provides some interesting information about the theory of languages. We recall that two Cauchy sequences $(s_n)_{n\in\omega}$ and $(t_n)_{n\in\omega}$ of words are equivalent if it turns out that, as n increases, the distance between s_n and t_n decreases. Therefore, we need a greater monoid to distinguish s_n and t_n.
The space $(\widehat{A^*}, d^*)$ is a complete, compact, zero-dimensional and totally disconnected metric space. The elements of $\widehat{A^*}$ are called profinite words and the ultrametric d^* represents the continuous extension of d to $\widehat{A^*}$. Because of this, we get that:

$$d^*(s,t) = \lim_{n\to\infty} d(s_n, t_n).$$

We get the following interesting characterisation of regular languages: a language is regular iff its closure in the space $(\widehat{A^*}, d^*)$ is a clopen (both open and closed) set. Other results concerning varieties of regular languages are summarised in the following. We recall that a class of regular languages is a variety if it is closed under the main Boolean operations on languages.

1. A class of languages is a variety iff it is axiomatised by a set of profinite equations. This characterisation can be expressed by using Priestly duality: the space $(\widehat{A^*}, d^*)$ corresponds to the dual space of the Boolean algebra of regular languages. Formally, it turns out that a set of regular languages on A^* is a lattice (and subsequently a Boolean algebra) iff it can be defined by a set of profinite equations;

2. String concatenation is dual operation to the residuals of right/left product on the Boolean algebra of regular languages. These operations are included in extended Priestly duality by other relational structures on dual space. More precisely, it results that the set of all possible regular languages on A^*, endowed with the aformentioned quotienting operations, corresponds to the dual of the topological monoid of profinite words. Thus, we get that:

    - The dual of the residual of the left/right product corresponds to the product of profinite words;
    - The closure of the Boolean operations on the Boolean algebra of regular languages corresponds to the left/right continuity of string concatenation;
    - The equational property corresponds to product associativity.

These results represent the starting point for other future developments in the theory of formal languages.

Chapter 1

Topology and set theory

In this chapter we report definitions about sets, relations, topology and metric spaces (with examples) which will be useful in order to understand many concepts involved in this thesis.

1.1 Sets and relations

Definition 1 (preorder). Let S be a set and $\leq \subseteq S \times S$ be a binary relation. A pair $(S, \leq)$ is a *preorder* if $\leq$ satisfies the reflexive and transitive properties.

A preorder $(S, \leq)$ is a **partial order** if it satisfies the antisimmetric property.
A partial order $(S, \leq)$ is a **total order** if, $\forall\ a, b \in S,\ a \leq b \vee b \leq a$.

Definition 2 (filter). Let S be a set. Then $\mathcal{F}$ is a *filter* on S if $\mathcal{F} \subseteq \mathcal{P}(S)$ and, $\forall\ T, V \subseteq S$, the following properties hold:

1. *upper closure*: $T \in \mathcal{F} \wedge T \subseteq U \Rightarrow U \in \mathcal{F}$;
2. $\cap$-*closure*: $T \in \mathcal{F} \wedge U \in \mathcal{F} \Rightarrow T \cap U \in \mathcal{F}$.

$\mathcal{F}$ is a **proper filter** if $\mathcal{F} \neq \mathcal{P}(S)$.
$\mathcal{F}$ is an **ultrafilter** if it is a proper filter and, $\forall\ X \subseteq S, X \in \mathcal{F} \vee (S \backslash X) \in \mathcal{F}$.
A proper filter $\mathcal{F}$ on S is **prime filter** if, $\forall\ X, Y \subseteq \mathcal{P}(S), X \cup Y \in \mathcal{F} \Rightarrow X \in \mathcal{F} \wedge\ Y \in \mathcal{F}$.

1.2 Topology

Definition 3 (topological space). A *topological space* is a pair (S, τ) where S is a set and τ is a family of subsets of S satisfying the following properties:

1. $\emptyset, S \in \tau$;
2. $\forall\ V \subseteq \tau, \bigcup V \in \tau$;
3. $\forall\ V \subseteq_f \tau, \bigcap V \in \tau$.

The elements of τ are called **open sets**. A set T is **closed** if $S \setminus T$ is open. A set which is both open and closed is called **clopen**.
A set I is a **neighborhood** of $x \in S$ if $\exists\ O \in \tau\ (x \in O \subseteq I)$.
If $X \subseteq S$, then the **closure** of X is the set $\overline{X} = \{x \in S \mid (\forall\ I \in \mathcal{N}_x)\ I \cap X \neq \emptyset\}$.

Example 1 Let $S = \{a, b, c\}, \tau_1 = \{\emptyset, S, \{a\}\}$ and $\tau_2 = \{\emptyset, S, \{a\}, \{b\}, \{a, b\}\}$. Then (S, τ_1) and (S, τ_2) are topological spaces.

Definition 4 (base). Let (S, τ) be a topological space. A family $\mathcal{B} \subseteq \tau$ is a *base* if $\forall\, O \in \tau\ \exists\, V \subseteq\ \mathcal{B}\ (O = \bigcup V)$.

Let (S, τ) be a topological space. A **cover** of $T \subseteq S$ is a family $\mathcal{F}$ of open sets such that: $T \subseteq \bigcup_{F \in \mathcal{F}} F$. A cover can be either finite or infinite (depending whether the number of sets is finite or not).
(S, τ) is **0-dimensional** if there exists a base $\mathcal{B}$ for (S, τ) such that, $\forall\, V \in \mathcal{B}, V$ is clopen.

Definition 5 (connected space). Let (S, τ) be a topological space. Then S is *connected* if it is not the union of two nonempty open sets O and P such that $O \cap P = \emptyset$.

A topological space that is not connected is called **disconnected**.

Definition 6 (continuous function). Let (S, τ_1), (T, τ_2) be topological spaces and $f : S \to T$ be a function. Then f is *continuous* if, $\forall\, O \in \tau_2, f^{-1}[O]$ is an open set in S.

Definition 7 (homeomorphism). Let (S, τ_1) and (T, τ_2) be topological spaces. Then $f : S \to T$ is a *homeomorphim* if f is continuous, bijective and the inverse map $f^{-1} : T \to S$ is continuous.

If there is a homeomorphism from (S, τ_1) to (T, τ_2) then (S, τ_1) and (T, τ_2) are called **homeomorphic**.
A topological space (X, τ) is $\mathbf{T_0}$ if, $\forall\, x, y \in X, x \neq y$, there exists an open set U such that either $(x \in U \wedge y \notin U)$ or $(y \in U \wedge x \notin U)$.
(X, τ) is $\mathbf{T_2}$ (or **Hausdorff**) if, $\forall\, x, y \in X, x \neq y$, $\exists\, U \in \mathcal{N}_x, V \in \mathcal{N}_y$ such that $U \cap V = \emptyset$. This situation is represented on Figure 1.1.

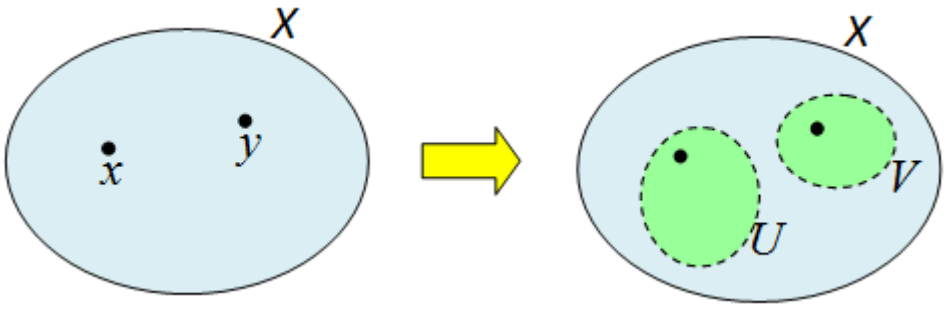


Figure 1.1: topological representation of the the T_2-axiom.

A subset X of a topological space (S, τ) is **dense** if $\overline{X} = S$.
A **sequence** $(s_n)_{n \in \omega}$ on S is a function $s : \omega \to S$.
A **subsequence** of a sequence $(s_n)_{n \in \omega}$ is a sequence $(s_{n_r})_{r \in \omega}$ where $n_r < n_k$ if $r < k$.
$x \in S$ is a **limit** of a sequence $(s_n)_{n \in \omega}$ if, $\forall\, I \in \mathcal{N}_x\ \exists\, N \in \omega$ such that $\forall\, n > N\ s_n \in I$. This is usually written as $\lim_{n \to \infty} (s_n) = x$.
A topological space (S, τ) is **compact** if every cover of S admits a finite subcover.
(S, τ) is **discrete** if, $\forall\, X \subseteq S, X$ is open.
(S, τ) is **Boolean** (or **Stone**) if it is T_0, compact and 0-dimensional.

1.3 Metric spaces

Definition 8 (metric space). A *metric space* is a pair (X, d) where X is a set and $d : X \times X \to \mathbb{R}^+$ is a function, called distance, that satisfies the following properties $\forall\ x, y, z \in X$:

- Identity: $d(x, y) = 0$ iff $x = y$;
- Simmetry: $d(x, y) = d(y, x)$;
- Triangular inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

Example 2 Every metric space is Hausdorff.

A metric space (X, d) is **ultrametric** if d satisfies the strong triangular inequality: $d(x, y) \leq max\{d(x, z), d(z, y)\}$.
Let $(s_n)_{n\in\omega}$ be a sequence defined on a metric space (X, d). Then (s_n) is **Cauchy sequence** if, $\forall\ \epsilon > 0\ \exists\ n_\epsilon > 0$ such that $\forall\ n, m > n_\epsilon\ d(s_n, s_m) < \epsilon$.
Let (X, d_X) and (Y, d_Y) be metric spaces and $f : X \to Y$ be a function. Then f is an **isometry** if, $\forall\ x, y \in X, d_X(x, y) = d_Y(f(x), f(y))$.

Example 3 (metric spaces). Euclidian spaces are the best known examples of metric spaces. Normed vector spaces (where $d(x, y) = \|x - y\|$) are other examples of metric spaces.

Proposition 1. *Let (X, d_X), (Y, d_Y) be metric spaces. Then $f : X \to Y$ is continuous if, for every $x \in X$ and $\epsilon > 0$, $\exists\ \delta > 0$ such that $\forall\ y \in X$, if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \epsilon$.*

Definition 9 (uniformly continuous function). Let (X, d_X), (Y, d_Y) be metric spaces and $f : X \to Y$ be a function. Then f is *uniformly continuous* if, $\forall\ \epsilon > 0\ \exists\ \delta > 0$ such that $\forall\ x, y \in X,$ if $d_X(x, y) < \delta$, then $d_Y(f(x), f(y)) < \epsilon$.

Definition 10 (complete metric space). Let (X, d) be a metric space. Then (X, d) is *complete* if every Cauchy sequence (s_n) defined on X admits a limit.

Definition 11 (completion of a metric space). Let (X, d_X) be a metric space and (Y, d_Y) be a complete metric space. (Y, d_Y) is a *completion* of (X, d_X) if there exists an isometry $f : X \to Y$ such that $f[X]$ is dense in Y.

It is well known that the completion of a metric space is unique up to isomorphism.
Let (X, d_X) be a metric space and (Y, d_Y) be a complete metric space. Every uniformly continuos function $f : X \to Y$ can be uniquely extended to a uniformly continuos function $\widehat{f} : \widehat{X} \to Y$, where $(\widehat{X}, \widehat{d_X})$ is the completion of (X, d_X). Let $(s_n)_{n\in\omega}$ be Cauchy sequence on X. Moreover, we define a sequence $\widehat{f}((s_n)_{n\in\omega})$ on Y as $\widehat{f}((s_n)_{n\in\omega}) = (f((s_n)_{n\in\omega}))$. We need to verify that it is a Cauchy sequence. Since f is uniformly continuous then, given $\epsilon > 0$, we can choose $\delta > 0$ such that, $\forall\ x, y \in X,$ if $d_X(x, y) < \delta$ then $d_Y(f(x), f(y)) < \epsilon$. So, given the same δ as above there exists n_δ such that, $\forall\ n, m > n_\delta$, $d(s_n, s_m) < \delta$. If we take n_δ then, $\forall\ n, m > n_\delta$, $d(f(s_n), f(s_m)) < \epsilon$. This means that $f(s_n)$ is a Cauchy sequence.

Chapter 2

Algebras

In this chapter the main concepts related to algebras are introduced. In addition some algebraic structures, which are recalled in this thesis, are described.

2.1 Algebraic notions

Let A be a nonempty set and n a positive integer. A map $f : A^n \to A$ is called a **n-ary operation** on A. An operation is a **constant** if $n = 0$.

Definition 12 (type of algebras). $\mathfrak{F}$ is a *type of algebras* (simply a *type*) if $\mathfrak{F}$ is a family of function symbols each with its arity.

$\mathfrak{F}_n \subseteq \mathfrak{F}$ is the set of all function symbols of $\mathfrak{F}$ of arity n.

Definition 13 ($\mathfrak{F}$-algebra). Let $\mathfrak{F}$ be a type. Then $\mathbf{A} = \langle A, f^{\mathbf{A}} \rangle_{f \in \mathfrak{F}}$ is an $\mathfrak{F}$-*algebra* if, for every $f \in \mathfrak{F}_n$, $f^{\mathbf{A}}$ is an n-ary operation on A.

An algebra $\mathbf{A}$ is **finite** if $|A|$ is finite and **trivial** if $|A| = 1$.
Given a tuple $\langle A, f_1, \ldots, f_k \rangle$, where $f_1, \ldots, f_k$ are operations on A, we say that $\langle A, f_1, \ldots, f_k \rangle$ is an **algebra of type** $(\mathbf{n_1}, \ \ldots \ , \mathbf{n_k})$ if $f_i : A^{n_i} \to A$ for every $i \leq k$.

Example 4 (lattice). An algebra $\langle R, \vee, \wedge \rangle$ of type $(2, 2)$ is a **lattice** if the following properties hold $\forall\ a, b, c \in R$:

1. *Commutative laws*:
   - $a \vee b = b \vee a$;
   - $a \wedge b = b \wedge a$;
2. *Associative laws*:
   - $a \vee (b \vee c) = (a \vee b) \vee c$;
   - $a \wedge (b \wedge c) = (a \wedge b) \wedge c$;
3. *Absorption laws*:
   - $a \vee (a \wedge b) = a$;
   - $a \wedge (a \vee b) = a$;
4. *Idempotent laws*:

- $a \vee a = a$;
- $a \wedge a = a$.

A lattice is **distributive** if the following laws hold:

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \text{ and } a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

An algebra $\langle R, \wedge \rangle$ ($\langle R, \vee \rangle$) of type (2) is a **meet-semilattice** (**join-semilattice**) if $\wedge$ ($\vee$) satisfies the associative, commutative and idempotent laws.
An algebra $\langle R, \wedge, \vee, 1, 0 \rangle$ of type $(2, 2, 0, 0)$ is a **bounded lattice** if $\langle R, \wedge, \vee \rangle$ is a lattice and the following properties, a.k.a. *identity laws*, hold $\forall\ a \in R$:

- $a \wedge 0 = 0$;
- $a \wedge 1 = a$.

Example 5 (Boolean algebra). A **Boolean algebra** is an algebra $\langle A, \vee, \wedge, ', 0, 1 \rangle$ of type $(2, 2, 1, 0, 0)$ such that:

1. $\langle A, \wedge, \vee, 0, 1 \rangle$ is a bounded distributive lattice;
2. $a \wedge a' = 0$, $a \vee a' = 1$.

Example 6

- Let S be a set, $A, B \subseteq S$ and $\cup, \cap, '$ be binary operations on $\mathcal{P}(S)$ defined as follows:
  - $A \cup B = \{x \mid x \in A \vee x \in B\}$;
  - $A \cap B = \{x \mid x \in A \wedge x \in B\}$;
  - $A' = \{x \in S \mid x \notin A\}$.

  Then $\mathcal{P}(\mathcal{S}) = \langle \mathcal{P}(S), \cup, \cap, ', \emptyset, S \rangle$ is a Boolean algebra;
- Let $2 = \{0, 1\}$, where $0, 1$ represent the truth values *false* and *true*. Then the algebra $\mathbf{2} = \langle 2, \wedge, \vee, ', 0, 1 \rangle$ of truth values is a Boolean algebra;
- Let (S, τ) be a topological space. Then $\langle \tau, \cup, \cap \rangle$ is a distributive lattice and $\langle Cl(S, \tau), \cup, \cap, ', \emptyset, S \rangle$ is a Boolean algebra.

Definition 14 (congruence). Let $\mathbf{A}$ be an algebra of type $\mathfrak{F}$ and θ be an equivalence relation on A. Then θ is a *congruence* on $\mathbf{A}$ if, $\forall\ f \in \mathfrak{F}_n$ and $\forall\ a_i, b_i \in A$:

$$a_1 \theta b_1, \ldots, a_n \theta b_n \Rightarrow f(a_1, \ldots, a_n) \theta f(b_1, \ldots, b_n).$$

$\mathbf{Con}(\mathbf{A})$ denotes the **lattice of congruences** on $\mathbf{A}$.
The following particular congruences can be defined on A:

- $\Delta = \{(a, a) \mid a \in A\}$ is the identity relation on A;
- $\nabla = \{(a, b) \mid a, b \in A\}$ is the total relation on A.

Δ and ∇ denote respectively the least and the greatest congruence on $\mathbf{Con}(\mathbf{A})$.

Definition 15 (homomorphism). Let $\mathbf{A}$ and $\mathbf{B}$ be $\mathfrak{F}$-algebras and $\phi : \mathbf{A} \to \mathbf{B}$ be a function. ϕ is a *homomophism* from $\mathbf{A}$ to $\mathbf{B}$ if, $\forall\ f \in \mathfrak{F}_n, a_1, \ldots, a_n \in A$:

$$\phi f^{\mathbf{A}}(a_1, \ldots, a_n) = f^{\mathbf{B}}(\phi a_1, \ldots, \phi a_n).$$

A homomorphism $\phi : \mathbf{A} \to \mathbf{B}$ is:

- An **endomorphism** if $\mathbf{A} = \mathbf{B}$;
- An **isomorphism** if ϕ is bijective.

$\mathbf{B}$ is a **homomorphic image** of $\mathbf{A}$ if there exists a surjective homomorphism $\phi : A \to B$.

Definition 16 (subalgebra). Let $\mathbf{A}$ and $\mathbf{B}$ be $\mathfrak{F}$-algebras. Then $\mathbf{B}$ is a *subalgebra* of $\mathbf{A}$, denoted by $\mathbf{B} \leq \mathbf{A}$, if $B \subseteq A$ and, $\forall\ f \in \mathfrak{F}_n$ and $b_1, \ldots, b_n \in B$:

$$f^{\mathbf{B}}(b_1, \ldots, b_n) = f^{\mathbf{A}}(b_1, \ldots, b_n).$$

Definition 17 (direct product). Let $\mathbf{A}$ and $\mathbf{B}$ be $\mathfrak{F}$-algebras. The *direct product* between $\mathbf{A}$ and $\mathbf{B}$ is an algebra whose universe is $A \times B$ and such that $\forall\ f \in \mathfrak{F}_n (1 \leq i \leq n), a_i \in A$ and $b_i \in B$:

$$f^{\mathbf{A}\times\mathbf{B}}(\langle a_1, b_1\rangle, \cdots, \langle a_n, b_n\rangle) = \langle f^{\mathbf{A}}(a_1, \cdots, a_n), f^{\mathbf{B}}(b_1, \cdots, b_n)\rangle.$$

The direct product between $\mathbf{A}$ and $\mathbf{B}$ is simply denoted by $\mathbf{A} \times \mathbf{B}$. The direct product can be easily generalized to a family of algebras $(\mathbf{A}_i)_{i\in I}$.
Let $(\mathbf{A_i})_{i\in I}$ be an indexed family of $\mathfrak{F}$-algebras. The **projection map** $\pi_j : \prod_{i\in I} A_i \to A_j$ is defined $\forall\ j \in I$ as $\pi_j(a) = a_j \in A_j$. This gives a surjective homomorphism $\pi_j : \prod_{i\in I} \mathbf{A_i} \to \mathbf{A_j}$.

Definition 18 (subdirect product). Let $\mathbf{A}$ be an algebra and $(\mathbf{A_i})_{i\in I}$ be an indexed family of algebras. Then $\mathbf{A}$ is a *subdirect product* of $\prod_{i\in I} \mathbf{A_i}$ if:

- $\mathbf{A} \leq \prod_{i\in I} \mathbf{A_i}$;
- $\forall\ i \in I\ \pi_i(A) = A_i$.

Definition 19 (directly indecomposable algebra). Let $\mathbf{A}$ be an algebra. $\mathbf{A}$ is *directly indecomposable* if $\mathbf{A} \ncong \mathbf{A_1} \times \mathbf{A_2}$, with $\mathbf{A_1}$ and $\mathbf{A_2}$ nontrivial algebras.

Let $\mathfrak{F}$ be a type and X be a set of distinct objects called **variables**. Then $T_{\mathfrak{F}}(X)$ (or $T(X)$ when there is no danger of confusion) is the **set of terms** of type $\mathfrak{F}$ over X. $T(X)$ is defined as the smallest set satisfying the following properties:

1. $X \cup \mathfrak{F}_0 \subseteq T(X)$;
2. $\forall\ p_1, \ldots, p_n \in T(X)$ and $\forall\ f \in \mathfrak{F}_n$, $f(p_1, \ldots, p_n) \in T(X)$.

$\mathbf{T}(X)$ is a **term algebra** if it is an $\mathfrak{F}$-algebra with universe $T_{\mathfrak{F}}(X)$ and $\forall\ f \in \mathfrak{F}_n, p_i \in T(X), 1 \leq i \leq n$:

$$f^{\mathbf{T}(X)} : \langle p_1, \cdots, p_n\rangle \mapsto f(p_1, \cdots, p_n).$$

Let $p(x_1, \cdots, x_n) \in T(X)$ and $\mathbf{A}$ be an $\mathfrak{F}$-algebra. Then $p^{\mathbf{A}} : A^n \to A$ is a mapping defined as follows $\forall\ a_1, a_2, \ldots, a_n \in A$:

1. If p is a variable x_i:
$$p^{\mathbf{A}}(a_1, \ldots, a_n) = a_i.$$
In other words, $p^{\mathbf{A}}$ is the i-th projection map;

2. If p is a term on the form $f(p_1(x_1,\ldots,x_n),\ldots,p_k(x_1,\ldots,x_n))$, with $f \in \mathfrak{F}_k$, then:

$$p^{\mathbf{A}}(a_1,\ldots,a_n) = f^{\mathbf{A}}(p_1^{\mathbf{A}}(a_1,\ldots,a_n),\ldots,p_k^{\mathbf{A}}(a_1,\ldots,a_n)).$$

An **identity** of type $\mathfrak{F}$ on a set of variables X is an expression $(p \approx q)$ where $p, q \in T(X)$.
An $\mathfrak{F}$-algebra **A satisfies an identity** $p(x_1,\ldots,x_n) \approx q(x_1,\ldots,x_n)$ if, $\forall\ a_1,\ldots,a_n \in A$:

$$p^{\mathbf{A}}(a_1,\ldots,a_n) = q^{\mathbf{A}}(a_1,\ldots,a_n).$$

This is commonly denoted by $\mathbf{A} \models (p \approx q)$.
A class K of algebras **satisfies a set of identities** Σ if, $\forall\ ((p \approx q) \in \Sigma, \mathbf{A} \in K)$, $\mathbf{A} \models (p \approx q)$.
If Σ is a set of identities, then $M(\Sigma)$ is the **class of algebras satisfying** Σ.

Definition 20 (equational class of algebras). A class K of algebras is an *equational class* if there exists a set of identities Σ such that $K = M(\Sigma)$.

It is said that K is **axiomatized** by Σ.
Let K be a class of $\mathfrak{F}$-algebras and $\mathbf{A} \in K$. Then:

- $\mathbf{A} \in \mathrm{I}(K)$ if there exists $\mathbf{B} \in K$ such that $\mathbf{A} \cong \mathbf{B}$;
- $\mathbf{A} \in \mathrm{S}(K)$ if there exists $\mathbf{B} \in K$ such that $\mathbf{A} \leq \mathbf{B}$;
- $\mathbf{A} \in \mathrm{IS}(K)$ if $\mathbf{A} \in \mathrm{I}(K)$ and $\mathbf{A} \in \mathrm{S}(K)$.

Definition 21 (free algebra). Let K be a family of $\mathfrak{F}$-algebras, X be a set of variables, $T(X)$ be the set of terms over X and θ be the congruence defined as follows: $\theta_K(X) = \bigcap \Phi_K(X)$, where $\Phi_K(X) = \{\phi \in \mathbf{Con}(\mathbf{T}(X)) \mid \mathbf{T}(X)/\phi \in \mathrm{IS}(K)\}$. Then $\mathbf{F_K}(X) = \mathbf{T}_{\mathfrak{F}}(X)/\theta_K(X)$ is the *K-free algebra* over X.

Definition 22 (variety of algebras). $\mathcal{V}$ is a *variety of algebras* if it is closed under subalgebras, homomorphic images and direct products.

Actually a variety can be defined also as the class of all algebraic structures of the same type satisfying a given set of identities. Birkhoff proved the equivalence of these definitions through the well known Birkhoff's theorem, a.k.a. **HSP theorem** (where the letters H, S and P refer to the operations of homomorphic image, subalgebra and product). The importance of this theorem is due to the capability of describing a class of structures through a set of equations.
The class of Boolean algebras is an example of variety.

2.2 Semigroups and monoids

Definition 23 (semigroup). An algebra $\langle S, \cdot\rangle$ of type (2) is a *semigroup* if $\cdot$ is associative.

Definition 24 (monoid). An algebra $\langle M, \cdot, 1\rangle$ of type $(2,0)$ is a *monoid* if $\langle M, \cdot\rangle$ is a semigroup and, $\forall\ a \in M$, $a \cdot 1 = 1 \cdot a = a$.

Every semigroup $\langle S, \cdot\rangle$ can be extended to a monoid $\langle S \cup \{1\}, \cdot, 1\rangle$ by defining $a \cdot 1 = 1 \cdot a = a$.

Definition 25 (ordered monoid). An *ordered monoid* is a quadruple $\langle M, \cdot, 1, \leq \rangle$ where $\langle M, \cdot, 1 \rangle$ is a monoid and $\leq$ is an order relation on M such that $\forall\ x, y, u, v \in M$:

$$x \leq y \Rightarrow uxv \leq uyv.$$

Definition 26 (free monoid). Let A be an alphabet, $\cdot$ be the string concatenation and ϵ be the empty string. Then $\mathbf{A}^* = \langle A^*, \cdot, \epsilon \rangle$ is the *free monoid* on A.

Chapter 3

Formal languages

In this chapter the main important concepts related to the theory of formal languages and the theory of automata are defined, in particular focusing on regular languages since they play an important role in this thesis.
Furthermore in the last part even some interesting theorems related to the aforementioned topics are reported.

3.1 Formal languages theory

The theory of formal languages is an important mathematical field studying strings and sets of strings (i.e. languages). The theory of formal languages has many applications in Linguistics, in Computer Science, etc. Subsequently we have many different approaches:

- **Generative approach**: a language is seen as the set of all possible strings generated by a grammar (see below);
- **Recognisable approach**: a language is seen as the set of all possible strings accepted by an automata;
- **Denotational approach**: a language is defined as denotation of an expression (e.g. a regular expression);
- **Algebraic approach**: a language is defined through its algebraic properties (see for example semigroup theory and Kleene algebras);
- **Transforming approach**: a language can be obtained by transforming another language, which is simpler to analyse than the starting one.

3.2 Chomsky classification

An **alphabet** is a finite set of symbols.
A **language** of alphabet A is a subset of the set of strings on A.
Formal grammars are useful to generate languages in a simple way. We can distinguish two different kinds of grammars:

- **Generative grammars**: most popular and widely adopted, they consist in a set of rules that are used to generate all possible strings belonging to a given language. Such strings are produced by rewriting a set of symbols according to the given rules;

- **Analytical grammars**: a string is provided as input to an automaton and subsequently analysed. Thus, the answer will be yes/no depending whether the input string belongs to the language or not. For example this approach is useful in order to describe a linguistic parser.

Noam Chomsky proposed a definition of generative grammar which is at the same time both concise and clear.

Definition 27 (grammar). A quadruple $\mathcal{G} = \langle V, A, P, S \rangle$ is a *grammar* if:

- V is a finite and nonempty set of non-terminal symbols. Its elements are denoted by capital letters $A, B, C, \ldots$. They are commonly called variables;
- A is a finite nonempty alphabet of terminal symbols. Its elements are denoted by lowercase letters $a, b, c, \ldots$. It is required that $A \cap V = \emptyset$;
- $S \in V$ is the initial symbol;
- P is a finite and nonempty set of production rules, where a production rule is a pair (α, β) with $\alpha \in (V \cup A)^+$ and $\beta \in (V \cup A)^*$ (see below). A production rule is denoted by $\alpha \to \beta$.

Let $\mathcal{G} = \langle V, A, P, S \rangle$ be a grammar.
A **one-step reduction** is defined as follows, for all $w, v \in (V \cup A)^*$:

$$w \to v \text{ iff } \exists\ s, t, x, y \in (V \cup A)^* \mid (w = sxt) \wedge (x \to y \in P) \wedge (v = syt).$$

It is said that v can be derived from w in one step.
$\to^*$ is the reflexive and transitive closure of $\to$.

Definition 28 (language generated by a grammar). Let $\mathcal{G}$ be a grammar. Then the *language generated by* $\mathcal{G}$ is defined as: $L(\mathcal{G}) = \{w \in A^* \mid S \to^* w\}$.

Example 7 (languages generated by a grammar).

- The language $L = \{a^n b^m c^n \mid n, m > 0\}$ can be generated by the grammar $\mathcal{G} = \langle V, A, P, S \rangle$ where:
  - $V = \{B, S\}$;
  - $A = \{a, b, c\}$;
  - $P = \{S \to aSc, S \to aBc, B \to Bb, B \to b\}$.

  For example the string $a^2b^3c^2$ is produced by the following reduction:

  $$S \to aSc \to a^2Bc^2 \to a^2Bbc^2 \to a^2Bb^2c^2 \to a^2b^3c^2;$$

- The language $L = \{a^{2n} b^m c^n \mid n \geq 0, m > 0\}$ can be generated by the grammar $\mathcal{G} = \langle V, A, P, S \rangle$ where:
  - $V = \{B, S\}$;
  - $A = \{a, b, c\}$;
  - $P = \{S \to a^2Sc, S \to B, B \to Bb, B \to b\}$.

  For example the string a^4bc^2 is produced by the following reduction:

  $$S \to a^2Sc \to a^4Sc^2 \to a^4Bc^2 \to a^4bc^2;$$

- The language $L = \{a^n b^n c^n \mid n \geq 0\}$ can be generated by the grammar $\mathcal{G} = \langle V, A, P, S \rangle$ where:
  - $V = \{A, B, C, S\}$;
  - $A = \{a, b, c\}$;
  - $P = \{S \to abc, S \to aSBc, cB \to AB, AB \to AC, AC \to BC, BC \to Bc, bB \to bb\}$.

  For example the string $a^2b^2c^2$ is produced by the following reduction:

  $$S \to aSBc \to a^2bcBc \to a^2bABc \to a^2bACc \to a^2bBCc \to a^2bBc^2 \to a^2b^2c^2.$$

Chomsky proposed a classification of generative grammars in four different types. This is the well known **Chomsky hierarchy**.
Let $\mathcal{G} = \langle V, A, P, S \rangle$ be a grammar. $\mathcal{G}$ is **right-regular** if its rules are on the form $B \to aC$ or $B \to a$, with $B, C \in V$ and $a \in A$. $\mathcal{G}$ is **left-regular** if C and a are inverted on the second part of the rule.
In the following the main features of the grammar types are described (w.r.t. the classification of figure 3.1):

- **Type-0 grammar** (*unlimited grammar*): this type includes all formal grammars. Since no restrictions are made on production rules **recursively enumerable languages** can be generated by them;
- **Type-1 grammar** (*context-sensitive grammar*): this type includes grammars having production rules on the form $\alpha B \beta \to \alpha \gamma \beta$, with $B \in V, \alpha, \gamma, \beta \in (V \cup A)^*$ and $\gamma \neq \epsilon$. The rule $B \to \epsilon$ is allowed if, $\forall\ (\alpha, \beta) \in P, B \notin \beta$.
  The languages generated by these grammars are called **context-sensitive languages**;
- **Type-2 grammar** (*context-free grammar*): this type includes grammars having production rules on the form $B \to \gamma$, with $B \in V$ and $\gamma \in (V \cup A)^*$.
  The languages generated by these grammars are called **context-free languages**;
- **Type-3 grammar** (*regular grammar*): this type includes grammars which are either right or left regular but not both them, otherwise the language is not regular. The rule $B \to \epsilon$ is allowed if, $\forall\ (\alpha, \beta) \in P, B \notin \beta$.
  The languages generated by these grammars are called **regular languages**.

All finite languages are of type 3.

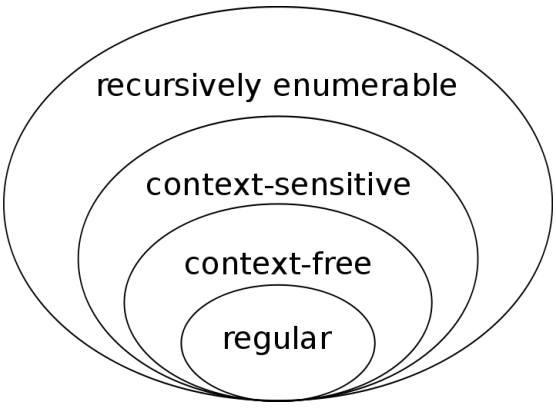


Figure 3.1: formal grammars representation according to Chomsky hierarchy.

Table 3.1 summarizes the previous grammars with their accepted languages and the corresponding minimal automaton that recognise them.

Type	Grammar	Language	Automaton
0-type	(unlimited)	Recursively enumerable	Turing machine
1-type	Context sensitive	Context sensitive	Linear automaton
2-type	Context free	Context free	ND pushdown automaton
3-type	Regular	Regular	Finite state

Table 3.1: formal grammars according to Chomsky hierarchy.

A context-free grammar $\mathcal{G}$ (or a type-2 grammar w.r.t. Chomsky hierarchy) is said to be in **Chomsky normal form** (shortly denoted by CNF) if, $\forall\, p \in P, p$ assumes one of these possible forms (where S is the starting symbol):

1. $A \to BC$;
2. $A \to a$;
3. $S \to \epsilon$.

3.3 Regular languages

Regular languages offer interesting properties. The compactness used to express them and their operations are important features so that they are widely involved in different applications (e.g. input parsing and programming language design).

3.3.1 Language operations

Let L_1 and L_2 be languages of a finite alphabet A. Then these are the main language operations that can be defined:

- **Concatenation**: $L_1 \cdot L_2 = \{uv \mid u \in L_1 \wedge v \in L_2\}$;
- **Kleene star**: $L_1^* = \bigcup_{i \in \omega} L_1^i$, where L_1^i can be recursively defined as follows:

$$L_1^0 = \{\epsilon\}, \ldots, L_1^{i+1} = \{w\alpha \mid w \in L_1^i \wedge \alpha \in L_1\};$$

- **Kleene plus**: $L_1^+ = L_1^* \backslash \{\epsilon\}$;
- **Intersection**: $L_1 \cap L_2 = \{u \mid u \in L_1 \wedge u \in L_2\}$;
- **Union**: $L_1 \cup L_2 = \{u \mid u \in L_1 \vee u \in L_2\}$;
- **Complement**: $L_1^C = \{u \mid u \in A^* \wedge u \notin L_1\}$;
- **Right quotient**: $L_1 \backslash L_2 = \{u \mid \exists\, v \in L_2 \mid uv \in L_1\}$;

- **Left quotient**: $L_1/L_2 = \{u \mid \exists\ v \in L_1 \mid uv \in L_2\}$;
- **Reversal**: let $u = a_1 \dots a_n$ be a string. The reversal of u, denoted by u^R, is defined as: $u^R = a_n \dots a_1$. Then $L_1^R = \{u^R \mid u \in L_1\}$;
- **Shuffle**: $L_1 \sqcup\!\sqcup L_2 = \{u = v_1 w_1 v_2 w_2 \dots v_n w_n \mid v_1, v_2, \dots, v_n \in L_1 \wedge w_1, w_2, \dots, w_n \in L_2\}$.

Languages are closed under $\cup$, $\cdot$ and *.

Example 8 (language operations). Let $A = \{a, b, c, d\}, L_1 = \{ab, a^2c, ad, ba\}$ and $L_2 = \{ad, bc, d\}$. Then:

- $L_1 \cdot L_2 = \{abad, ab^2c, abd, a^2cad, a^2cbc, a^2cd, adad, adbc, ad^2, ba^2d, babc, bad\}$;
- $L_1 \cap L_2 = \{ad\}$;
- $L_1 \cup L_2 = \{ab, a^2c, ad, ba, bc, d\}$;
-
  - $L_1^C = \{\epsilon, a, b, c, d, a^2, ac, b^2, bc, bd, a^3, a^2b, a^2d, \dots\}$;
  - $L_2^C = \{\epsilon, a, b, c, a^2, ab, ac, ba, b^2, bd, a^3, a^2b, a^2c, a^2d, \dots\}$;
-
  - $L_1 \backslash L_2 = \{\epsilon, a\}$;
-
  - $L_1/L_2 = \{\epsilon\}$;
-
  - $L_1^* = \{\epsilon, ab, a^2c, ad, ba, abab, aba^2c, abad, adba, ababab, \dots\}$;
  - $L_2^* = \{\epsilon, ad, bc, d, adad, adbc, ad^2, adadad, \dots\}$;
-
  - $L_1^R = \{ba, ca^2, da, ab\}$;
  - $L_2^R = \{da, cb, d\}$;
-
  - $L_1 \sqcup\!\sqcup L_2 = \{ababdab^2c, a^2cadbad, babcba, \dots\}$;
  - $L_2 \sqcup\!\sqcup L_1 = \{adada^2cd, a^2cbcbad, ba^2dab^2c, \dots\}$.

3.3.2 Regular expressions

Let A be an alphabet. The set of **regular expressions** (shortly **regex**) is defined by induction as follows:

- $\underline{\emptyset}, \underline{\epsilon}$ and $\underline{a}$ $(a \in A)$ are regex;
- if r, s are regex then r^*, rs and $r + s$ are regex.

The semantics of a regular expression is the following:

- $|\underline{\emptyset}| = \emptyset$;
- $|\underline{\epsilon}| = \{\epsilon\}$;
- $|\underline{a}| = \{a\}$;
- $|r^*| = |r|^*$;
- $|rs| = |r| \cdot |s|$;
- $|r + s| = |r| \cup |s|$.

A regex is very useful to express a language in a compact way.

Example 9 (regular expressions).

- $|a^*(bb)^*| = \{\epsilon, ab^2, ab^4, a^2b^2, a^3b^2, \dots\}$;
- $|aa^* + ab + bb| = \{a, ab, b^2, a^2, a^3, \dots\}$;
- $|(ac)^*(ab)(ab)^*| = \{ab, acab, acabab, acacababab, acacacab, \dots\}$.

3.4 Finite automata and regular languages

Definition 29 (deterministic finite automaton). A quintuple $\mathcal{A} = \langle S, A, \lambda, s_0, F\rangle$ is a *deterministic finite automaton* if:

- S is the nonempty set of states;
- A is the nonempty set of input alphabet;
- $\lambda : S \times A \to S$ is the transition function;
- $s_0 \in S$ is the initial state;
- $F \subseteq S$ is the nonempty set of final states.

For every $a \in A$ let $\lambda_a : S \to S$ be the function defined by $\lambda_a(s) = \lambda(s, a)$. We define the **extended transition function** $\lambda_x : S \to S$ $(x \in A^*)$ by induction as follows:

$$\lambda_x(s) = \begin{cases} s & \text{if } x = \epsilon \\ \lambda_a(\lambda_y(s)) & \text{if } x = ya. \end{cases}$$

The composition operation between two extended transition functions λ_x and λ_y can be characterised as follows:

$$\lambda_y \circ \lambda_x = \lambda_{xy}.$$

Definition 30 (recognised language). Let $\mathcal{A} = \langle S, A, \lambda, s_0, F\rangle$ be a finite automaton and $w = x_1 \dots x_n$ be a word on A. Then $L(\mathcal{A})$, the *language recognised by* $\mathcal{A}$, is the set $L(\mathcal{A}) = \{w \in A^* \mid \lambda_w(s_0) \in F\}$.

Definition 31 (transition monoid). Let $\mathcal{A} = \langle S, A, \lambda, s_0, F\rangle$ be a deterministic finite automaton and $\lambda_x : S \to S$ $(x \in A^*)$ be the extended transition function on $\mathcal{A}$. Then $\mathbf{TM}_\mathcal{A} = \langle \{\lambda_x\}_{x \in A^*}, \circ, \lambda_\epsilon\rangle$ is called the *transition monoid* of $\mathcal{A}$.

The transition monoid is finite and it determines a monoid homomorphism defined as follows:

$$\mathbf{A}^* \to \mathbf{TM}_{\mathcal{A}}$$
$$x \mapsto \lambda_x.$$

This homomorphism determines a congruence on A^* which partitions A^* in a finite number of equivalence classes. This congruence is defined as follows, $\forall\ x, y \in A^*$:

$$x \equiv_{\mathcal{A}} y \text{ iff } \lambda_x = \lambda_y.$$

This means that the strings x and y are indistinguishable by $\mathcal{A}$. So, $L(\mathcal{A})$, the language recognised by $\mathcal{A}$, is finite union of equivalence classes.

Theorem 1. *Let L be a language. Then the following statements are equivalent:*

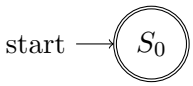
1. *L is regular;*
2. *There exists a regular expression E such that $|E| = L$;*
3. *L is recognised by a finite automaton.*

Corollary 1. *(Kleene theorem). The class of regular languages is the smallest class containing finite languages which is closed under union, concatenation and Kleene star.*

Example 10 (deterministic finite automaton for regular languages).

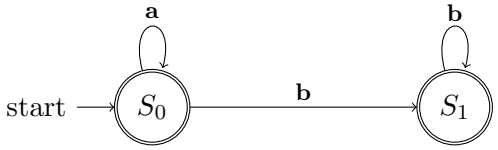
- $L = \{a^n b^n \mid n \geq 0\}$ is not regular. It turns out immediatly that it is not possible to build a finite automaton to accept L. Behind this impossibility there is the following intuition: a finite automaton has no memory. Because of this it cannot "remember" the exact number of a and subsequently recognise the same number of b;
- $L = \{\epsilon\} = \emptyset^*$ (the empty string language) is regular. In this special case the corresponding automaton that accepts L is trivially composed by a single starting/final state without transitions, as shown on Figure 3.2;

Figure 3.2: finite automaton for regular language $L = \emptyset^*$.



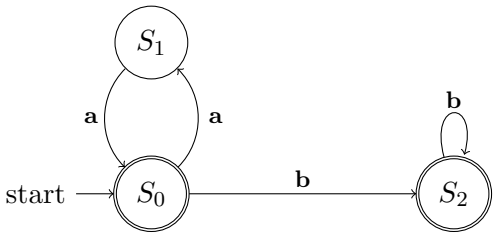
- $L = \{a^n b^m |\ n, m \geq 0\}$ is regular. A finite automaton that recognises L is shown on Figure 3.3;

Figure 3.3: finite automaton for regular language $L = \{a^n b^m |\ n, m \geq 0\}$.



- $L = \{a^{2n}b^m |\ n, m \geq 0\}$ is regular. The automaton shown on Figure 3.4 is able to recognise L.

  Figure 3.4: finite automaton for regular language $L = \{a^{2n}b^m |\ n, m \geq 0\}$.



3.5 Recognisable sets

Let $\langle M, \cdot, 1\rangle$ be a monoid and $P \subseteq M$. Then the relation $\sim_P$ on M, defined by:

$$s \sim_P t \text{ iff } (\forall\ x, y \in M(xsy \in P \text{ iff } xty \in P)),$$

is a congruence, called the **syntactic congruence** of P.
The quotient $M/\sim_P$ determines the **syntactic monoid** of P and $\eta_P : M \to M/\sim_P$ is the **syntactic homomorphism** of P. Notice that $M/\sim_P$ may be infinite, but $P = m_1/\sim_P \cup \cdots \cup m_k/\sim_P$ for some $m_1, \ldots, m_k \in M$.

Definition 32 (recognisable subset). Let $\mathbf{M} = \langle M, \cdot, e\rangle$ be a monoid, $S \subseteq M$ and θ be a congruence defined on M of finite index. Then S is a *recognisable subset* of $\mathbf{M}$ if $S = m_1/\theta \cup \cdots \cup m_k/\theta$ for some $m_1, \ldots, m_k \in M$.

It follows that a subset P of a monoid $\mathbf{M} = \langle M, \cdot, e\rangle$ is a recognisable set of $\mathbf{M}$ iff the syntactic monoid $M/\sim_P$ is finite.
The following lemma characterises recognisable subsets.

Lemma 1. *Let $\mathbf{M} = \langle M, \cdot, e\rangle$ be a monoid and $S \subseteq M$. Then S is recognisable iff there exist a finite monoid $\mathbf{N} = \langle N, \cdot, e\rangle$, a homomorphism $\varphi : \mathbf{M} \to \mathbf{N}$ and $P \subseteq N$ such that $S = \varphi^{-1}[P]$.*

Proof.

- $\Rightarrow$: if S is recognisable then, according to Definition 32, there exists a congruence θ of finite index on M such that $S = m_1/\theta \cup \cdots \cup m_k/\theta$ for some $m_1, \ldots, m_k \in M$. Therefore, $m_1/\theta \cup \cdots \cup m_k/\theta$ is finite and the homomorphism $\varphi : \mathbf{M} \to \mathbf{M}/\theta$ proves the implication;

- $\Leftarrow$: let $\langle N, \cdot, e\rangle$ be a finite monoid, $\varphi : \mathbf{M} \to \mathbf{N}$ be a homomorphism and $P \subseteq N$ such that $S = \varphi^{-1}[P]$. Then the congruence $Ker(\varphi)$ associated to φ is of finite index. Moreover, S is union of $Ker(\varphi)$-classes since it is counterimage of $P \subseteq N$. Thus, according to Definition 32, S is recognisable.

□

Example 11 (recognisable subsets). Let $\mathcal{A} = \langle S, A, \lambda, s_0, F\rangle$ be a deterministic automaton and $P = \{\lambda_x \mid \lambda_x(s_0) \in F\}$. Then $L(\mathcal{A})$ is a recognisable subset of A^* (according to Definition 32) because there exists a homomorphism $\varphi : \mathbf{A}^* \to \mathbf{TM}_{\mathcal{A}}$ and $L(\mathcal{A}) = \varphi^{-1}[P]$.

Proposition 2. *Let L be a recognisable subset of A^*, $\mathbf{M} = \langle M, \cdot, e\rangle$ be a finite monoid, $P \subseteq M$ and $\varphi : \mathbf{A}^* \to \mathbf{M}$ be a homomorphism such that $L = \varphi^{-1}[P]$. Then the finite automaton $\mathcal{A} = \langle M, A, \lambda, e \in M, P\rangle$, where $\lambda(m, a) = \varphi(a) \cdot m$ for every $m \in M$ and $a \in A$, recognises L. Moreover, the transition monoid of $\mathcal{A}$ is isomorphic to the monoid $\mathbf{M}$.*

Proof. It is possible to prove by induction that $\lambda_w(m) = \varphi(w) \cdot m$, for every $w \in A^*$ and $m \in M$. Then $w \in L$ iff $\lambda_w(e) = \varphi(w) \in P$. □

By Proposition 2 it follows that if L is a recognisable subset of A^* then there exists a finite automaton that recognises L (this automaton in general may not be unique).
Now we determine a finite monoid that recognises L starting from a finite automaton that recognises it. Among all finite monoids recognising L there is one which has minimal cardinality and we are interested on it. This is exaclty the syntactic monoid.
At this point it can be said that a language L is union of congruence classes of $\sim_L$ but this congruence is not necessarily of finite index. The next theorem characterises L even when it is finite union of equivalence classes.

Theorem 2. *(MyHill). Let $\langle A^*, \cdot, \epsilon\rangle$ be a free monoid and $L \subseteq A^*$ be a language. Then the following statements are equivalent:*

1. *L is regular;*
2. *L is a recognisable subset of A^*;*
3. *The syntactic monoid of L is finite and it is a monoid of minimal cardinality which recognises L.*

Proof.

- (1) ⇒ (2): if L is regular then, by Theorem 1, there exists a finite automaton that recognises L. The transition monoid of this automaton is a finite monoid which recognises L;
- (2) ⇒ (3): since L is a recognisable subset of A^* then, by Definition 32, there exists a congruence θ on A^* of finite index such that $L = a_1/\theta \cup \cdots \cup a_k/\theta$ for some $a_1, \ldots, a_k \in A^*$. We know that, $\forall\ u, v \in A^*, u \sim_L v$ iff $(\forall\ x, y \in A^*(xuy \in L$ iff $xvy \in L))$. Moreover, since θ is a congruence then, $\forall\ u, v, x, y \in A^*(u\theta v \Rightarrow xuy\theta xvy)$, so that $xuy \in L$ and $xvy \in L$. This means that, $\forall\ u, v \in A^*(u\theta v \Rightarrow (\forall\ x, y \in A^*(xuy \in L$ iff $xvy \in L)))$ or, equivalently, $u\theta v \Rightarrow u \sim_L v$. Therefore, $\theta \subseteq \sim_L$ and subsequently the syntactic monoid of L is finite;
- (3) ⇒ (1): by Proposition 2.

□

By (2) ⇒ (3) of Theorem 2 it results that if the syntactic monoid is finite then it has minimal cardinality among the ones recognising the language L. Therefore, the automaton associated to the finite syntactic monoid of L is the minimal automaton recognising L.

Proposition 3. *The syntactic monoid of a recognisable language is isomorphic to the transition monoid of its minimal automaton.*

Proof. Let L be a recognisable language of A^*, $\mathcal{A} = \langle S, A, \lambda, s_0, F\rangle$ be the minimal automaton that accepts L, $\mathbf{TM}_{\mathcal{A}} = \langle \{\lambda_x\}_{x\in A^*}, \circ, \lambda_\epsilon\rangle$ be the transition monoid of $\mathcal{A}$ and $\sim_L$ be the syntactic congruence on L. We need to verify that, $\forall\ u, v \in A^*$, $u \sim_L v$ iff $\lambda_u = \lambda_v$:

- $\Rightarrow$: let $u \sim_L v$ and $s \in S$. For every $s \in S, s$ can be reached from s_0. Therefore, there exists $w \in A^*$ such that $\lambda_w(s_0) = s_1$. Then, since $u \sim_L v$, $\forall\ y \in A^*$:

  $$wuy \in L \text{ iff } wvy \in L.$$

  This can be also written as:

  $$\lambda_{wuy}(s_0) \in F \text{ iff } \lambda_{wvy}(s_0) \in F$$

  or:

  $$\lambda_y(\lambda_u(s_1)) \in F \text{ iff } \lambda_y(\lambda_v(s_1)) \in F.$$

  This means that $\lambda_u(s_1) = \lambda_v(s_1)$ and therefore u and v define the same transition on $\mathbf{M}$;

- $\Leftarrow$: let $u, v \in A^*$ be two words that define the same transition on $\mathbf{M}$. So, if $wuy \in L$ then $\lambda_{wuy}(s_0) \in F$. We can also say that wuy and wvy define the same transition on $\mathbf{M}$, so we get that $\lambda_{wvy}(s_0) \in F$ and subsequently $wvy \in L$. This proves that $wuy \sim_L wvy$.

□

Chapter 4

Profinite topology and duality

This chapter describes, after giving some basic definitions, the main relations that exist among regular languages, Boolean algebras and profinite words. The background that you need to understand them is part of the previous chapters.

4.1 Profinite topology and profinite words

Let A be a finite alphabet, $u, v \in A^*$, $\langle M, \cdot, e\rangle$ be a finite monoid and $\varphi : \mathbf{A}^* \to \mathbf{M}$ be a homomorphism. Then φ separates u and v if $\varphi(u) \neq \varphi(v)$. It can be also said by extension that $\langle M, \cdot, e\rangle$ separates u and v if there is a homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}$ that separates them.

Proposition 4. *For every distinct $u, v \in A^*$ there exists a finite monoid $\langle M, \cdot, e\rangle$ that separates u and v.*

Proof. Let $u, v \in A^*$ and $u \neq v$. The language $\{u\}$ is recognisable. Therefore, according to Lemma 1, there exists a homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}$ (with $\langle M, \cdot, e\rangle$ finite monoid) that recognises $\{u\}$. This means that $\varphi^{-1}[\varphi(u)] = \{u\}$. It follows that $\varphi(u) \neq \varphi(v)$ and therefore φ separates u and v. □

Now let $r(u, v) = \min\{|M| \mid \langle M, \cdot, e\rangle \text{ separates } u \text{ and } v\}$. Then $d : A^* \times A^* \to \mathbb{R}^+, d(u, v) = 2^{-r(u,v)}$, is a distance (see below) and (A^*, d) is a metric space (the assumptions $\min \emptyset = \infty$ and $2^{-\infty} = 0$ are done).

Proposition 5. *d satisfies the following properties $\forall\ u, v, w \in A^*$:*

1. *d is an ultrametric, so:*
   - 1.1. *$d(u, v) = 0$ iff $u = v$;*
   - 1.2. *$d(u, v) = d(v, u)$;*
   - 1.3. *$d(u, w) \leq \max\{d(u, v), d(v, w)\}$;*
2. *$d(uw, vw) \leq d(u, v)$ and $d(wu, wv) \leq d(u, v)$.*

Proof.

1.1
- $\Rightarrow$: let $d(u, v) = 0$. Then it must be $r(u, v) = \infty$ and this means that we need an infinite monoid to separate u and v. Now suppose that $u \neq v$. Then, according to Proposition 4, there exists a finite monoid that separates u and v. This is absurd since we have previously proved that such monoid is infinite and therefore $u = v$;

- $\Leftarrow$: let $u = v$. Then u and v cannot be separated by a finite monoid. Therefore, since, by assumption, $\min \emptyset = \infty$ and $2^{-\infty} = 0$, it follows that $d(u, v) = 0$;

1.2 $d(u, v) = d(v, u)$ because a homomorphism φ that separates u and v does not take into account the order of the words (in both cases $\varphi(u) \neq \varphi(v)$);

1.3 Let $\langle M, \cdot, e\rangle$ be a monoid separating u and w, so that $d(u, w) > 0$. Then $\langle M, \cdot, e\rangle$ separates either u and v or v and w. Therefore, $\min\{r(u, v), r(v, w)\} \leq r(u, w)$ and this means that $d(u, w) \leq \max\{d(u, v), d(v, w)\}$;

2. If a finite monoid separates uw and vw of course it distinguishes also u and v. It follows that $d(uw, vw) \leq d(u, v)$ and, by (1.2), $d(uw, vw) \leq d(v, u)$.

□

Proposition 6. *The topology induced by the ultrametric d on A^* is the discrete topology.*

Proof. Let $u \in A^*$, $\mathbf{M} = \langle M, \cdot, e\rangle$ be the syntactic monoid of $\{u\}$ and n be the cardinality of M. For every $v \in A^*, v \neq u$, $\mathbf{M}$ separates u and v. So, $r(u, v) \leq n$ and, by this, it results $2^{-n} \leq 2^{-r(u,v)}$. By the generality of v it derives that the ball $\mathcal{B}(u, 2^{-n})$ contains only $\{u\}$ and hence every singleton of A^* is open. So the topology induced by d on A^* is the discrete topology. □

A sequence $(s_n)_{n\in\omega}$ of words is a **Cauchy sequence on** A^* if, $\forall\ \epsilon > 0$, there exists k such that, $\forall\ n, m > k, d(s_n, s_m) < \epsilon$. The idea behind this definition is that as n, m increase, the distance between two words s_n and s_m decreases (i.e. it becomes harder to distinguish s_n and s_m) and therefore the cardinality of the monoid we need to separate them increases.

The completion of (A^*, d), called the **profinite topological space** and denoted by $(\widehat{A^*}, d^*)$, is complete. The elements of $\widehat{A^*}$ are called **profinite words** on A.

$(\widehat{A^*}, d^*)$ can be constructed as follows. Let C be the set of Cauchy sequences that can be defined on A^*. The equivalence relation $\sim$ can be defined on C, $\forall\ (s = (s_n)_{n\in\omega}, t = (t_n)_{n\in\omega})$ as:

$$s \sim t \text{ iff the sequence } (s_0, t_0, \ldots, s_n, t_n) \text{ is a Cauchy sequence.}$$

$\widehat{A^*}$ can be defined as the set of equivalence classes of C module $\sim$.

The metric $d : A^* \times A^* \to \mathbb{R}^+$ is a uniformly continuous function from the metric space $(A^* \times A^*, d')$, where $d'((u, v), (w, z)) = max\{d(u, w), d(v, z)\}$, into the euclidean metric space $\mathbb{R}^+$. The metric d on A^* can be extended to a metric d^* on $\widehat{A^*}$ as follows:

$$d^*(s, t) = \lim_{n\to\infty} d(s_n, t_n).$$

Lemma 2. *If $x \in A^*$ then $\{x\}$ is open in $(\widehat{A^*}, d^*)$.*

Proof. Let $\mathbf{M}$ be the syntactic monoid of $\{x\}$ with $n = |M|$. If if $t = (t_n)_{t\in\omega}$ is a Cauchy sequence of words such that $d^*(x, t) < 2^{-n}$, then there exists $k > 0$ such that $\forall\ n > k, d(x, t_n) < 2^{-n}$. This implies $t_n = x$ for every $n > k$. □

Lemma 3. *If $L \subseteq A^*$ then $L = \overline{L} \cap A^*$.*

Proof. If $x \in A^*$ and $x \in \overline{L}$ then, since $\{x\}$ is open in $(\widehat{A^*}, d^*)$, $\{x\} \cap L \neq \emptyset$. This means that $x \in L$ and therefore $L = \overline{L} \cap A^*$. □

Theorem 3. *$(\widehat{A^*}, d^*)$ is compact.*

Proof. $(\widehat{A^*}, d^*)$ is complete and therefore we need to prove that , $\forall\ n > 0$, (A^*, d) is covered by a finite number of balls $\mathcal{B}(x, r)$ such that $r < 2^{-n}$. Let $\sim_n$ be a congruence on A^* defined as follows, $\forall\ u, v \in A^*$:

$$u \sim_n v \text{ iff } \varphi(u) = \varphi(v) \text{ for every homomorphism } \varphi : \mathbf{A}^* \to \mathbf{M},$$
$$\text{where } \mathbf{M} = \langle M, \cdot, e \rangle \text{ is a finite monoid and } |M| \leq n.$$

Since A is finite then there is a finite number of such homomorphisms and therefore $\sim_n$ is a congruence of finite index. Moreover, $d(u, v) < 2^{-n}$ iff u and v cannot be separeted by $\mathbf{M}$ or, in other words, $u \sim_n v$. It derives that the $\sim_n$-classes are the open balls $\mathcal{B}(x, r)$, with $r < 2^{-n}$, and they cover A^*. □

Another important consideration w.r.t. A^* regards the continuity of the string concatenation, as shown in the next proposition.

Proposition 7. *The string concatenation is uniformly continuous from $A^* \times A^*$ to A^*.*

Proof. By exploiting first property (1.3) and then (2) of Proposition 5 it turns out that, $\forall\ u, w, v, z \in A^*$:

$$d(uv, wz) \leq \max\{d(uv, uz), d(uz, wz)\} \leq \max\{d(v, z), d(u, w)\}.$$

□

Since the string concatenation on A^* is uniformly continuous we can extend it to $\widehat{A^*}$ as shown in the following proposition.

Proposition 8. *Every homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}$ (where $\mathbf{M} = \langle M, \cdot, e \rangle$ is a discrete and finite monoid) is uniformly continuous and can be uniquely extended to a uniformly continuous homomorphism $\widehat{\varphi} : \widehat{\mathbf{A}^*} \to \mathbf{M}$.*

Proof. Let $\langle M, \cdot, e \rangle$ be a discrete, finite monoid and $\varphi : \mathbf{A}^* \to \mathbf{M}$ be a homomorphism. For every $u, v \in A^*$ such that $d(u, v) < 2^{-|M|}$, φ does not separate u and v because $\varphi(u) = \varphi(v)$. By extension even $\mathbf{M}$ does not separate u and v. This means that φ is a uniformly continuous function and we can uniquely extend it to its completion (see Chapter 1) by the homomorphism $\widehat{\varphi} : \widehat{\mathbf{A}^*} \to \mathbf{M}$. □

The importance of the profinite topological space $(\widehat{A^*}, d^*)$ is explained in the following proposition.

Proposition 9. *The profinite topology induced by d^* on $\widehat{A^*}$ is the least topology that makes every homomorphism $\widehat{\varphi} : \widehat{\mathbf{A}^*} \to \mathbf{M}$ ($\mathbf{M} = \langle M, \cdot, e \rangle$ finite monoid) continuous.*

Proof. We suppose, considering Proposition 8, to have a topological space $(\widehat{A^*}, \tau)$ such that τ makes every homomorphism from A^* to a finite monoid $\langle M, \cdot, e \rangle$ continuous. We have to prove that, $(\forall\ \epsilon > 0, x \in \widehat{A^*})$, the ball $\mathcal{B}(x, \epsilon) = \{y \in \widehat{A^*} \mid d(x, y) < \epsilon\}$ is open in τ.

Let $n = \min\{m \in \omega \mid 2^{-m} < \epsilon\}$. So, $d(x,y) < \epsilon$ iff there not exists a monoid $\langle M, \cdot, e\rangle$ separating x and y having $|M| < n$. Let $\Phi = \{\varphi : A^* \to M \mid |M| < n\}$. Then $\mathcal{B}(x,\epsilon) = \bigcap_{\varphi\in\Phi} \varphi^{-1}[\varphi(x)]$. Since τ is a topology every set $\varphi^{-1}[\varphi(x)]$ is open and since A is finite then Φ is finite. It derives that $\mathcal{B}(x,\epsilon)$ is open and this proves the proposition. □

$\langle \widehat{A^*}, \widehat{\cdot}, \widehat{\epsilon}\rangle$ is the **free profinite monoid on $\mathbf{A}^*$**, where:

- $\widehat{\cdot}$ is the product of profinite words which in fact represents the extension of $\cdot$ to $\widehat{A^*}$;
- $\widehat{\epsilon}$ is the empty profinite word.

By Proposition 8 it derives that every homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}$ can be uniquely extended to a homomorphism $\widehat{\varphi} : \widehat{\mathbf{A}^*} \to \mathbf{M}$. Thus, $\widehat{\varphi}$ separates $u, v \in \widehat{A^*}$ if $\widehat{\varphi}(u) \neq \widehat{\varphi}(v)$. Therefore, d^* can be even defined as follows (the assumptions $\min \emptyset = \infty$ and $2^{-\infty} = 0$ must still be valid):

$$d^*(u,v) = 2^{-r^*(u,v)}, \text{ where } r^*(u,v) = \min\{|M| \mid \langle M, \cdot, e\rangle \text{ separates } u \text{ and } v\}.$$

It is important to notice that $d^*(u,v) \in \{2^{-i} \mid i \in \omega\}$ (i.e. d^* is denumerable).
A **sequence of profinite words** $(s_n)_{n\in\omega}$ converges to a profinite word u iff, for every homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}, \hat{\varphi}(s_n) = \hat{\varphi}(u)$. Thus, every profinite word is limit of a Cauchy sequence of words.

Example 12 (profinite words). u^ω is an example of profinite word which is not word. More precisely: $u^\omega = \lim_{n\to\infty} u^{n!}$. This means that, $\forall\, u \in \widehat{A^*}$, $u^{n!}$ is a Cauchy sequence of words which admits a limit on $\widehat{A^*}$.

There is a final important result regarding the dimensionality and disconnection of $(\widehat{A^*}, d^*)$. We need the following lemma in order to prove the next theorem.

Lemma 4. *Let (X,d) be a metric space where $d : X \times X \to \mathbb{R}^+$ is a distance such that its image $Im(d)$ is denumerable. Then X is zero dimensional.*

Proof. We have to prove that the set $\mathcal{U} = \{\mathcal{B}(a,r) \mid a \in X \wedge r \notin Im(d)\}$ is a base of clopen.
The ball $\mathcal{B}(a,r)$, where $r \notin Im(d)$, is an open set by the definition of metric space. Now let $\mathcal{B}(a,t)$ be an open ball with $t \in Im(d)$. For any $x \in \mathcal{B}(a,t)$, there exists a ball $\mathcal{B}(x,k)$ such that $k < \dfrac{t - d(a,x)}{2}, k \notin Im(d)$ and $\mathcal{B}(x,k) \subset \mathcal{B}(a,t)$. This means that $\mathcal{U}$ is a base of X.
For any ball $\mathcal{B}(a,r) \in \mathcal{U}$, if $x \in \partial\mathcal{B}(a,r)$ then $d(x,a) = r$. This is absurd because $r \notin Im(d)$.This implies that if we set $Y = X \setminus \mathcal{B}(a,r)$ then $X = \overset{\circ}{\mathcal{B}}(a,r) \cup \overset{\circ}{Y}$. However, $\overset{\circ}{Y} = X \setminus \overset{\circ}{\mathcal{B}}(a,r) = X \setminus \mathcal{B}(a,r) = Y$ so Y is open and $\mathcal{B}(a,r)$ closed.

□

Theorem 4. *$(\widehat{A^*}, d^*)$ is 0-dimensional and totally disconnected.*

Proof. Since d^* satisfies the hypotheses of Lemma 4 then $(\widehat{A^*}, d^*)$ is 0-dimensional. Now it is well know by a theorem that a 0-dimensional and metric space is totally disconnected. □

Because $(\widehat{A^*}, d^*)$ is a metric, compact, 0-dimensional and totally disconnected space it is a Boolean space.

4.2 Connections to languages

It is important to characterise the main relations between the languages of words on A^* and the ones on $\widehat{A^*}$.
This first proposition shows a strong connection between recognisable and clopen subsets.

Proposition 10. *Let $P \subseteq \widehat{A^*}$. Then P is clopen iff P is a recognisable subset and its syntactic homomorphism $\widehat{\eta} : \widehat{\mathbf{A}^*} \to \mathbf{M}$ is continuous.*

Proof.

- $\Rightarrow$: according to the definition of syntactic congruence we get that:

$$\sim_P = \bigcap_{u,v \in \widehat{A^*}} ((u^{-1}Pv^{-1}) \times (u^{-1}Pv^{-1}) \cup (u^{-1}P^Cv^{-1}) \times (u^{-1}P^Cv^{-1})).$$

  Then if P is clopen each set $u^{-1}Pv^{-1}$ is also clopen.
  Now let (s_n, t_n) be a sequence, where $s_n, t_n \in (\sim_P)^C$, converging to (s, t). Then $s_n \nsim_P t_n$ and therefore there exist u_n, v_n profinite words such that $u_n s_n v_n \in P$ and $u_n t_n v_n \notin P$. (u_n, v_n) is a sequence which has a convergent subsequence since $\widehat{A^*} \times \widehat{A^*}$ is compact. Let (u, v) be its limit. Reminding that P and P^C are closed and the string concatenation is continuous on $\widehat{A^*}$ it turns out that $usv \in P$ and $utv \notin P$. Thus, $s \nsim_P t$: this means that $(\sim_P)^C$ is closed and $\sim_P$ is clopen. Since $\sim_P$ is clopen then, for every $s \in \widehat{A^*}$, there exists $O \in \mathcal{N}_s$ such that $O \times O \subseteq \sim_P$. Thus, O is contained in the $\sim_P$-classes of s and this means that the $\sim_P$-classes form an open partition of $\widehat{A^*}$. Now by compactness this partion is finite and then, according to Definition 32, we get that P is recognisable;

- $\Leftarrow$: let $\widehat{\eta} : \widehat{\mathbf{A}^*} \to \mathbf{M}$ be the syntactic homomorphism of P. Since each $\sim_P$-class is open then the syntactic homomorphism of P is continuous and since P is recognisable then M is finite. If we consider $P = \widehat{\eta}^{-1}[\widehat{\eta}[P]]$ then $\widehat{\eta}[P]$ is clopen in M because M is finite. Thanks to the continuity of $\widehat{\eta}$ we end up that P is clopen.

□

The following proposition relates regular languages with clopen subsets.

Proposition 11. *Let L be a language of A^*. Then L is regular iff $\overline{L}$ is clopen in $\widehat{A^*}$.*

Proof.

- $\Rightarrow$: if L is regular then, by Theorem 2, it is also a recognisable subset of A^*. So, there exists $\varphi : \mathbf{A}^* \to \mathbf{M}$ surjective homomorphism such that $L = \varphi^{-1}[\varphi[L]]$. Now we set $K = \hat{\varphi}^{-1}[\varphi[L]]$. So, since M is discrete then $\varphi[L] \subseteq M$ is clopen. The same can be said for K since $\hat{\varphi}^{-1}$ is continuous. It turns out that $\varphi = \hat{\varphi}$ on A^* and thus $L = \varphi^{-1}[\varphi[L]] \cap A^* = K \cap A^*$. Since we know that K is clopen and (A^*, d) is dense in $(\widehat{A^*}, d^*)$ then $K \cap A^*$ is dense in K. Hence we can write: $\overline{L} = \overline{K \cap A^*} = \overline{K} = K$. This means that $\overline{L}$ is clopen in $\widehat{A^*}$;

- $\Leftarrow$: since $\overline{L}$ is clopen in $\widehat{A^*}$ then, by Proposition 10, it results directly that L is a recognisable subset of $\widehat{A^*}$. Now if $\overline{L}$ is recognisable then there exists a homomorphism $\widehat{\eta} : \widehat{\mathbf{A}^*} \to \mathbf{M}$ with $\mathbf{M} = \langle M, \cdot, e \rangle$ finite monoid. Now let $P = \widehat{\eta}[\overline{L}]$ and η be nothing but $\widehat{\eta}$ restricted to A^*. So, we get $\overline{L} \cap A^* = \widehat{\eta}^{-1}[P] \cap A^* = \eta^{-1}[P]$. This means that L is recognisable and then, by Theorem 2, it is also regular.

□

This other proposition establishes important relations among regular languages and syntactic homomorhpisms.

Proposition 12. *Let L be a regular language of A^* and $u \in \widehat{A^*}$. Then the following statements are equivalent:*

(1) $u \in \overline{L}$;

(2) For every homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}, \hat{\varphi}(u) \in \varphi[L]$, where $\hat{\varphi}$ is the continuous extension of φ to $\widehat{A^*}$;

(3) There exists a homomorphism $\varphi : \mathbf{A}^* \to \mathbf{M}$ such that $\mathbf{M}$ recognises L and $\hat{\varphi}(u) \in \varphi[L]$;

(4) $\hat{\eta}(u) \in \eta[L]$, where η is the syntactic homomorphism of L and $\hat{\eta}$ is its continuous extension to $\widehat{A^*}$.

Proof.

- (1) $\Rightarrow$ (2): let $\varphi : \mathbf{A}^* \to \mathbf{M}$ be a homomorphism with $\mathbf{M} = \langle M, \cdot, e \rangle$ finite monoid. Then since $\hat{\varphi}$ is continuous, $\hat{\varphi}(\overline{L}) \subset \overline{\hat{\varphi}[L]}$ and $\overline{\hat{\varphi}[L]} = \hat{\varphi}[L] = \varphi[L]$ because $\mathbf{M}$ is discrete. Therefore, if $u \in \overline{L}$ then $\hat{\varphi}(u) \in \varphi[L]$;
- (2) $\Rightarrow$ (4): it derives immediatly by (1) and (2);
- (4) $\Rightarrow$ (3): it derives immediatly by (2) and (4);
- (3) $\Rightarrow$ (1): let $\varphi : \mathbf{A}^* \to \mathbf{M}$ be a homomorphism with $\langle M, \cdot, e \rangle$ finite monoid. Now let $(u_n)_{n \in \omega}$ be a sequence of words in A^* that admits a limit u. Thus, since $\hat{\varphi}$ is continuous then $\hat{\varphi}(u_n)$ converges to $\hat{\varphi}(u)$. On the other hand $\mathbf{M}$ is discrete and hence for n large enough it turns out that $\hat{\varphi}(u_n) = \hat{\varphi}(u)$. By (3) we get $\varphi(u_n) = \hat{\varphi}(u_n) \in \varphi[L]$. Finally since φ recognises L then we get $u_n \in \varphi^{-1}[\varphi[L]] = L$. So, it results that $u \in \overline{L}$.

□

4.3 Duality of Boolean algebras

The following theorem shows how the closure operators offer a nice and interesting behaviour w.r.t. the Boolean operations $\cap, \cup$ and ^{C} ($\mathbf{Reg}(A^*)$ and $\mathbf{Clopen}(\widehat{A^*})$ denote the Boolean algebras respectively of regular languages on A^* and clopen sets of $\widehat{A^*}$).

Theorem 5. *The maps $L \mapsto \overline{L}$ and $K \mapsto K \cap A^*$ can be seen as mutually inverse isomorphisms between $\mathbf{Reg}(A^*)$ and $\mathbf{Clopen}(\widehat{A^*})$. In particular for every $L_1, L_2 \in Reg(A^*)$ the following properties hold:*

1. $\overline{L_1^C} = (\overline{L_1})^C$;
2. $\overline{L_1 \cup L_2} = \overline{L_1} \cup \overline{L_2}$;
3. $\overline{L_1 \cap L_2} = \overline{L_1} \cap \overline{L_2}$.

Proof.

1. $\overline{L_1^C} = (\overline{L_1})^C$: by recalling Proposition 12 let η be the syntactic homomorphism of L_1. Therefore, since we have $L_1 = \eta^{-1}[\eta[L_1]]$ and $L_1^C = \eta^{-1}[\eta[L_1]^C]$ it results $\eta[L_1^C] = [\eta[L_1]]^C$ and the following chain of equalities derives:

$$\overline{L_1^C} = \hat{\eta}^{-1}[\eta[L_1^C]] = \hat{\eta}^{-1}[[\eta[L_1]]^C] = (\hat{\eta}^{-1}[\eta[L_1]])^C = (\overline{L_1})^C;$$

2. It is true since $\overline{L_1}$ and $\overline{L_2}$ are clopen in $\widehat{A^*}$;
3. It derives from 1 and 2.

This implies that the function from $\mathbf{Reg}(A^*)$ to $\mathbf{Clopen}(\widehat{A^*})$ which maps L into $\overline{L}$ is a homomorphism. It is bijective because the function from $\mathbf{Clopen}(\widehat{A^*})$ to $\mathbf{Reg}(A^*)$ which maps K into $K \cap A^*$ is its inverse. □

$\mathbf{Reg}(A^*)$ and $\mathbf{Clopen}(\widehat{A^*})$ behave as well as with basic Boolean algebras operations even with the operations left/right quotient and inverse homomorphism. This is shown in the following proposition.

Proposition 13. *Let $L \in Reg(A^*)$. Then for every $x, y \in A^*, \overline{x^{-1}Ly^{-1}} = x^{-1}\overline{L}y^{-1}$.*

Proof. Let η be the syntactic homomorphism of L. Since η recognises $x^{-1}Ly^{-1}$ then, by Proposition 12, we can derive the following implications:

$$\begin{aligned}
&u \in \overline{x^{-1}Ly^{-1}} \text{ iff}\\
&\hat{\eta}(u) \in \eta[x^{-1}Ly^{-1}] \text{ iff}\\
&\eta(x)\hat{\eta}(u)\eta(y) \in \eta[L] \text{ iff}\\
&\hat{\eta}(xuy) \in \eta[L] \text{ iff}\\
&xuy \in \overline{L} \text{ iff}\\
&u \in x^{-1}\overline{L}y^{-1}.
\end{aligned}$$

□

Definition 33 (lattice of languages). A *lattice of languages* is a class of regular languages closed under $\bigcap_f$ and $\bigcup_f$.

Definition 34 (variety of regular languages). A *variety* of regular languages is a class of regular languages closed under $\bigcap_f$, $\bigcup_f$, ^{C}, inverse image of the homomorphism between free monoids, left and right quotients.

Many attempts have been done to relax the definition of variety of languages in terms of closure of the required operations. Some of them are briefly introduced on [4, p. 2]. By recalling Definition 22 it is possible to define a variety of languages also by equations. Thus, the "point" created by this double definition establishes an important connection between Boolean algebras and profinite words which can be expressed in terms of

duality.
W.r.t. distributive lattices two main kinds of dualities can be used in order to express these connections.
Let D be a bounded distributive lattice. In **Stone duality**, D is based on S_D, the set of prime filters of D. Now, let $e(d) = \{\mathcal{F} \mid \mathcal{F} \text{ is a prime filter of } D \text{ and } d \in \mathcal{F}\}$. Birkhoff proved that there exists an embedding $e : D \hookrightarrow \mathcal{P}(S_D)$. If e is used then the resulting space is compact and 0-dimensional. This is maybe the most useful advantage when considering another structure on lattices and spaces such as the semigroup operation.
Priestley duality offers a new topological characterisation of the range of e. If one generates a topology τ not including only the sets in the range of e but also their complements then the dual space of the free Boolean extension of the lattice is obtained. Furthermore, if even the inclusion order on the space of prime filters is considered then it may be possible to reconstruct the original lattice. Thus, with Priestley duality the dual of a distributive lattice consists on the ordered topological space $(S_D, \subseteq, \tau)$, which is compact and totally order disconnected. This is the duality used in [4].
It can be proved that the dual space of $\mathrm{Reg}(A^*)$ is nothing but $\widehat{A^*}$. The following theorem summarizes this result.

Theorem 6. *The topological space underlying* $\widehat{\mathbf{A}^*}$ *is equal to the dual space of* $\mathbf{Reg}(A^*)$. *Furthermore, the canonical embedding is given by the topological closure* $e(L) = \overline{L}$.

Proof. Let $\mathbf{B_2}$ be the two-elements $\{0, 1\}$ Boolean algebra. We want to prove that $\mathbf{Reg}(A^*)^d$ is homeomorphic to $\widehat{\mathbf{A}^*}$ and $\widehat{\mathbf{A}^*}^d \cong \mathbf{Reg}(A^*)$. Hence, we have to check whether there exists a map $\varphi : \mathbf{Reg}(A^*) \times \widehat{\mathbf{A}^*} \to \mathbf{B_2}$ such that the following conditions hold:

1. For every $L \in Rec(A^*), \mu : \widehat{\mathbf{A}^*} \to \mathbf{B_2}$, defined by $\mu(t) = \varphi(L, t)$, is continuous;
2. For every $\mu : \widehat{\mathbf{A}^*} \to \mathbf{B_2}$, there exists $L \in Rec(A^*)$ such that $\mu(t) = \varphi(L, t)$;
3. For every $t \in \widehat{A^*}, \lambda : \mathbf{Reg}(A^*) \to \mathbf{B_2}$, defined by $\lambda[L] = \varphi(L, t)$, is a homomorphism;
4. For every $\lambda : \mathbf{Reg}(A^*) \to \mathbf{B_2}$ there exists $t \in \widehat{A^*}$ such that $\lambda[L] = \varphi(L, t)$.

We can define φ as follows:

$$\varphi(L, t) = \begin{cases} 1 & \text{if } t \in \overline{L} \\ 0 & \text{otherwise} \end{cases}$$

1. Let $L \in Rec(A^*)$. We can define $\mu(t) = \varphi(L, t)$. Since L is regular then, by recalling Proposition 11, $\overline{L}$ is clopen in $\widehat{A^*}$. This means that both $\overline{L}$ and $(\overline{L})^C$ are open in $\widehat{A^*}$. Therefore, for all $B \subseteq B_2, \mu^{-1}[B]$ is open in $\widehat{A^*}$ and thus μ is continuous;
2. Let $\mu : \widehat{\mathbf{A}^*} \to \mathbf{B_2}$ be a continuous mapping. So, both $\mu^{-1}(1)$ and $\mu^{-1}(0)$ are open. It turns out that $\mu^{-1}(1)$ is a clopen in $\widehat{A^*}$. If we take $L = \mu^{-1}(1) \cap A^+$ it results $\overline{L} = \mu^{-1}(1)$ and this means that $\overline{L}$ is clopen in $\widehat{A^*}$. Therefore, L is regular;
3. Let $t \in \widehat{A^*}$. We define $\lambda[L] = \varphi(L, t)$. We have to show that λ is a homomorphism. Therefore, since all Boolean operations can be expressed with $\cup$ and ^{C} we have to prove that for every L, M regular languages, $\lambda[L \cup M] = \lambda[L] \cup \lambda[M]$ and

$\lambda[A^+\backslash L] = \lambda[(L)^C]$. Since $\overline{\lambda[L \cup M]} = \overline{\lambda[L]} \cup \overline{\lambda[M]}$ it results $\lambda[L \cup M] = \lambda[L] \cup \lambda[M]$ by the definitions of λ and φ. Again, since L is regular then $\overline{L}$ is clopen in $\widehat{A^*}$ according to Proposition 11. So, it turns out that $\widehat{A^*}\backslash\overline{L} = \overline{A^+\backslash L}$ and, by the definitions of λ and φ, it results $\lambda[A^+\backslash L] = (\lambda[L])^C$;

4. Let $\lambda : \mathbf{Reg}(A^*) \to \mathbf{B_2}$ be a homomorphism. The filter $\mathcal{F} = \lambda^{-1}(1)$ is a maximal filter in $\mathbf{Reg}(A^*)$. According to this definition it is possible to state that:

   i) $L \in \mathcal{F} \wedge M \in Rec(A^*) \Rightarrow L \cup M \in \mathcal{F}$;
   ii) $L \in \mathcal{F} \wedge M \in \mathcal{F} \Rightarrow L \cap M \in \mathcal{F}$;
   iii) For every $L \in Rec(A^*)$, $L \in \mathcal{F} \dot{\vee} \overline{L} \in \mathcal{F}$ (not both them).

By $i), ii)$ and $iii)$ it turns out that the collection $\mathcal{F}' = \{\overline{L} \mid L \in F\}$ is an ultrafilter basis on $\widehat{A^*}$. Now since $(\widehat{A^*}, d^*)$ is a compact and Hausdorff space then $\bigcap \mathcal{F}' = \{t\}$, with $t \in \widehat{A^*}$.
The last verification to do is that, for every $L \in A^+, \lambda[L] = \varphi(L, t)$. $\lambda[L] = 1 \Rightarrow L \in \mathcal{F}$ and $\overline{L} \in \mathcal{F}$. This means that $\varphi(L, t) = 1$ according to the definition of φ. On the same way we get $\lambda[L] = 0 \Rightarrow L^C \in \mathcal{F}$ and $\overline{L^C} \in \mathcal{F}$. This means that $\varphi(L, t) = 0$ according to the definition of φ and this completes the proof. □

We define an **explicit equation** as a pair (u, v) of words such that $u, v \in A^*$.
A **profinite equation** is a pair (u, v) of words such that $u, v \in \widehat{A^*}$. In the special case that an identity is established between u and v the equation is usually called **profinite identity**.
A regular language $L \in A^*$ **satisfies** an equation $u \to v$ if $u \in \overline{L} \Rightarrow v \in \overline{L}$. It may happen to have a double implication so that $u \leftrightarrow v$ iff $u \in \overline{L} \Rightarrow v \in \overline{L} \wedge v \in \overline{L} \Rightarrow u \in \overline{L}$.
A **set of regular languages** $\mathcal{L}$ of A^* is defined by a set E of equations if, $\forall\ L \in \mathcal{L}$ and $\forall\ e \in E, L \models e$. The following theorem shows an important consequence of this.

Theorem 7. *A set of regular languages on A^* is a lattice of languages iff it can be defined by a set of profinite equations.*

Proof. Let D be a lattice of regular languages. We can define a dually quotient map $q_D : \widehat{A^*} \twoheadrightarrow S_D$ by $\mathcal{F}_u \mapsto \mathcal{F}_u \cap D$. Moreover, we can define this quotient map even by the preorder Q_D on $\widehat{A^*}$ as uQ_Dv iff $Q_D(\mathcal{F}_u) \subseteq Q_D(\mathcal{F}_v)$. This latter condition is nothing but requiring that for every $L \in D, u \in \overline{L} \Rightarrow v \in \overline{L}$.
Thus the preorder Q on $\widehat{A^*}$ which determines the quotient dual to D is exactly the equational theory of D. This quotient, denoted by Q_D, can be synthesized as follows:

$$Q_D = \{(u, v) \mid \forall\ L \in D, L \models u \to v\}.$$

In terms of duality, given a preorder Q on $\widehat{A^*}$ that arises a Priestley quotient $\widehat{A^*}/Q$, the corresponding lattice is the set $\{L \in \mathrm{Reg}(A^*) \mid \overline{L}$ is saturated w.r.t. the preorder $Q\}$. This means that, $\forall\ (u, v) \in Q$, $u \in L \Rightarrow v \in L$. However, as we have seen so far, this is exactly the set of languages defined by Q identifying each pair $(u, v) \in Q$ with the corresponding equation $u \to v$. Since coming from D, going to the preorder Q_D and then coming back to the set of languages defined by Q_D under duality gives us back D, it results that D is the set of languages defined by Q_D. □

Thus, we can now provide an equational description of the Boolean algebras of languages.

Corollary 2. *A set of regular language on A^* is a Boolean algebra of languages iff it can be defined by a set of equations on the form $u \leftrightarrow v$ $(u, v \in \widehat{A^*})$.*

4.4 Duality of regular languages

The second important result is that the string concatenation on $\widehat{A^*}$ is dual operation to the residual of the left/right product on $\mathbf{Reg}(A^*)$.
These operations are defined as follows, $\forall\ L, M, N \in Reg(A^*)$:

- **Right residual** of N by M: $M\backslash N = \{u \in A^* \mid Mu \subseteq N\} = \{u \in A^* \mid \forall\ v \in M, vu \in N\}$;
- **Left residual** of N by M: $N/M = \{u \in A^* \mid uM \subseteq N\} = \{u \in A^* \mid \forall\ v \in M, uv \in N\}$.

More precisely these additional operations are captured in extended Priestly duality by additional relational structures on the dual space.
In this case we get the algebra $\langle \mathrm{Reg}(A^*), \backslash, / \rangle$ of type $(2, 2)$, where the dual relation common to the two additional operations is functional and corresponds to the product of profinite words.
Further important considerations arise from the following theorem.

Theorem 8. *The dual space of the algebra $\langle Reg(A^*), \backslash, / \rangle$ under the extended duality is the topological monoid of profinite words $\langle \widehat{A^*}, \tau, \widehat{\cdot} \rangle$, where τ is a topology defined on $\widehat{A^*}$ and $\widehat{\cdot}$ is the concatenation of profinite words. The relational dual of the operations $\backslash$ and $/$ corresponds to $\widehat{\cdot}$. The closure of $Reg(A^*)$ under $\backslash$ and $/$ accounts for the right and left continuity of the string concatenation, respectively, and the equational property $(H\backslash K)/L = H\backslash(K/L)$ of $\langle Reg(A^*), \backslash, / \rangle$ corresponds to the associativity of the product.*

The proof of the previous theorem is very complicated and indeed it is not reported by the authors in [4] since it requires many advanced concepts related to duality theory. Nevertheless, part of its consequences are worth to be mentioned.
First, the syntactic ordered monoid of a regular language is nothing but the dual space of the subalgebra $\langle \mathrm{Reg}(A^*), \backslash, / \rangle$ generated starting from the singleton $\{L\}$ by lattice and residual operations. Moreover, the closure under product of $\langle \mathrm{Reg}(A^*), \backslash, / \rangle$ corresponds to the open mapping of the product of profinite words.
Second, if A is a finite alphabet then, for every $a \in A$, we can consider the residuals having denominator $\{a\}$. Thus, the quotienting operations can be denoted by $a^{-1}(\ )$ and $(\ )a^{-1}$ instead of $\{a\}\backslash(\ )$ and $(\ )\backslash\{a\}$.
Finally, a lattice of languages is called **quotienting algebra of languages** if it is closed under quotienting operations. It is interesting to characterise such lattices. So, if $u, v \in \widehat{A^*}$ then $L \models u \leq v$ if, $\forall\ x, y \in \widehat{A^*}, L \models xvy \to xuy$. In spite of this there exists a better way to characterise them by using the syntactic ordered monoid of L as described in the following proposition.

Proposition 14. *Let $L \in Reg(A^*)$ and $\mathbf{M} = \langle M, \cdot, e, \leq_L \rangle$ be its syntactic ordered monoid. Moreover, let $\eta : \mathbf{A}^* \to \mathbf{M}$ be its syntactic homomorphism. Then $L \models u \leq v$ iff $\hat{\eta}(u) \leq_L \hat{\eta}(v)$.*

Proof. $L \models u \leq v$ iff $\forall\ x, y \in A^*, \hat{\eta}(xvy) \in \eta[L] \Rightarrow \hat{\eta}(xuy) \in \eta[L]$. Since η is surjective and $\hat{\eta}(xvy) = \hat{\eta}(x)\hat{\eta}(v)\hat{\eta}(y)$ then the aformentioned implication is the same of the following one: $\forall\ s, t \in M, s\hat{\eta}(v)t \in \eta[L] \Rightarrow s\hat{\eta}(u)t \in \eta[L]$. This means that $\hat{\eta}(u) \leq_L \hat{\eta}(v)$. □

By Proposition 14 the following one derives directly.

Proposition 15. *Let $L \in A^*$ and η be its syntactic homomorphism. Then $L \models (u = v)$ iff $\hat{\eta}(u) = \hat{\eta}(v)$.*

So, an equational description of Boolean algebras of languages closed under quotients can be given in the next proposition.

Proposition 16. *Let $L_1, \ldots, L_n$ be languages of A^*. Then $\mathbf{L} = \{L_1, \ldots, L_n\}$ is a Boolean quotiening algebra iff it is possible to define it by a set of semigroup equations $u = v$ $(u, v \in \widehat{A^*})$.*

Although an equational description of lattice of languages has been previously given, now it may be useful to further generalize the obtained results. In fact, it is clear how in Theorem 7 the equations are constrained by the chosen alphabet. This constraint disappears under proper conditions that are explained in the following.
A **$\mathcal{C}$-homomorphism** is defined as a length-preserving homomorphism (w.r.t. the length of the words). So, let $\mathcal{C}$ be the class of free monoids containing $\mathcal{C}$-homomorphisms and closed under composition. Thus, a class of language lattices $\mathcal{L}$ associates to a finite alphabet A a lattice of languages denoted by $\mathcal{L}(A^*)$. Let $u \to v$ be an equation of $\mathcal{L}(A^*)$ and $\varphi : \mathbf{A}^* \to \mathbf{B}^*$ be a $\mathcal{C}$-homomorphism. Moreover, let $\hat{\varphi}(u) \to \hat{\varphi}(v)$ be an equation of $\mathcal{L}(B^*)$ and Σ be a countable alphabet. For every $u, v \in \widehat{\Sigma^*}, u \leq v$ is called **$\mathcal{C}$-identity**. Then a regular languge $L \in A^* \models u \leq v$ if, for every $\mathcal{C}-$homomorphism $\varphi : \Sigma^* \to A^*$, $L \models \hat{\varphi}(u) \to \hat{\varphi}(v)$.
In addition, it can be shown that a class of $\mathcal{C}$-homomorphisms is closed under quotiening and inverses of $\mathcal{C}$-homomorphisms iff it can be defined by a set of $\mathcal{C}$-identities $u \leq v$. Of course, similar results hold even for other identities like $u \leftrightarrow v, u \leq v$ or $u = v$. Essentially, a $\mathcal{C}$-identity can be seen as an equation in which each letter represents a variable.
Actually, it is important to point out that, although with the current analysis we have restricted our attention to the $\mathcal{C}$-class of length-preserving morphisms, these variables can be replaced by letters according to the kind of morphisms which are currently considered. For example, if we deal with the class of length-multiplying homomorphisms they can be replaced by words with the same length.

Chapter 5

Conclusions

In this master thesis we have studied a characterisation of regular languages from an algebraic point of view. We have seen how, starting from the common properties of regular languages like recognition by a finite automaton and closure under Boolean operations, a new topological characterisation is obtained. Furthermore, we have analysed the most important relations existing between Boolean algebras of regular languages and the clopen sets of topological spaces. In particular we have noticed how these results can be expressed in terms of duality w.r.t. some operations of regular languages. Although regular languages are fundamental in Computer Science, it may be interesting to focus also upon other kinds of languages (e.g. context-sensitive and context-free ones). It could be interesting to verify whether the mentioned duality relation still holds. In addition, we think that it is possible to extend the above results to other algebraic structures such as Church algebras (see [15] for more details).

Acknowledgements

This master thesis respresents the result of many years of studies, starting from Bachelor Degree and ending with Master Degree. Students and professors I met helped me in enforcing my interests in Computer Science and Mathematics and, subsequently, to extend my knowledge. For these reasons, I would like to thank a large amount of people.
First of all, I wish to thank the relator of this master thesis, professor Antonino Salibra, for accepting to guide me on this final and important task. A great contribution to this job is due also to PhD student Giordano Favro and researcher Alberto Carraro. Their help was fundamental for many proofs involved in this thesis.
I am particulary grateful to professors Marcello Pelillo, Agostino Cortesi and Nicoletta Cocco for the competence they offered during their lectures.
During my studies an important role has been played even by the colleagues of the Computer Science department. I appreciate them so much for the support to study we offered each other. So, I cannot forget persons like Mauro Tempesta, Marco Fiorucci, Antonio Panzetta, Alberto Meggiato, Gianluca Erboso, Enrico Eugenio, Lorenzo Brutti, Martina Olliaro, Gianni Burei and Andrea Lazzarotto. I will never forget even other people of Bachelor Degree, who I shared much time with, like Alessandro Tolomio, Luca Zorzi, Marco Stefan, Marco Gezzele and Vito Capacchione.
Even my referee colleagues are worth to be mentioned for many experiences we enjoyed in the free time. Therefore, I give credit to Alessio Giuliante, Stefano Fabbian, Enrico Galante, Michele Faggin, Huu Chinh Nguyen and others.
Last (but not least important) a special thank is dedicated to my family members. They have always encouraged and supported me in difficult decisions both related to studies and not.

List of abbreviations

Words

a.k.a.	also known as
e.g.	example given
etc.	and the rest
i.e.	that is
w.r.t.	with respect to

Logic

$(p \approx q)$	p is equal to q
$\Rightarrow$	implies
$\dot{\vee}$	xor
$\wedge$	and
$\vee$	or

Sets and relations

$A \subset_f B$	A is a finite subset of B with $A \neq B$
$A \subseteq B$	A is a subset of B
$A \subseteq_f B$	A is a finite subset of B
$A \,/\, \sim$	quotient of the relation $\sim$ defined on A
A^n	set of n-tuples of elements of A
$Cl(A, \tau)$	set of clopens of the topological space (A, τ)
$\bigcap$	arbitrary intersection
$\bigcup$	arbitrary union
$\emptyset$	empty set
$\exists!$	there exists a unique

$\|A\|$	cardinality of the set A
$\|w\|$	length of the string w
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}^+$	set of non negative real numbers
$\mathbb{R}$	set of real numbers
$\mathcal{B}(x,r)$	ball of center x and radius r
$\mathcal{N}_x$	set of neighbourhoods of the point x
$\mathcal{P}(A)$	set of subsets of A
$\bigcap_f$	finite intersection
$\bigcup_f$	finite union
ω	set of natural numbers
$\overline{A}$	closure of A
∂A	boundary of A
$\mathring{A}$	interior of A
$\|$	such that

Algebra

$\mathbf{A} \cong \mathbf{B}$	$\mathbf{A}$ is isomorphic to $\mathbf{B}$
$\mathbf{A} \models \phi$	$\mathbf{A}$ satisfies the formula ϕ
$\mathbf{A} \ncong \mathbf{B}$	$\mathbf{A}$ is not isomorphic to $\mathbf{B}$
$f^{\mathbf{A}}$	an operation f on the algebra $\mathbf{A}$

Functions

A^d	dual of A
$Ker(f)$	kernel of the function f
$f : A \hookrightarrow B$	f is an embedding from A to B
$f : A \to B$	f is a function from A to B
$f[A] = \{f(x) \mid x \in A\}$	image of A w.r.t. the function f
$f^{-1}[A] = \{f^{-1}(x) \mid x \in A\}$	inverse image of A w.r.t. the function f
$g \circ f$	composition of f and g
$n!$	factorial of n
$x \mapsto f(x)$	transformation law of the function f

Automata

$x \equiv_{\mathcal{A}} y$	strings x and y are equivalent w.r.t. automaton $\mathcal{A}$
D	Deterministic
ND	Non Deterministic

References

[1] S. Burris and H. P. Sankappanavar. *A course in universal algebra*. Vol. 78. Graduate Texts in Mathematics. New York: Springer-Verlag, 1981, pp. xvi+276. ISBN: 0-387-90578-2.

[2] Z. Ésik. "Axiomatizing the equational theory of regular tree languages". In: *The Journal of Logic and Algebraic Programming* 79.2 (2010). ISSN: 1567-8326.

[3] V. Geffert, C. Mereghetti, and B. Palano. "More concise representation of regular languages by automata and regular expressions". In: *Information and Computation* 208.4 (2010). ISSN: 0890-5401.

[4] M. Gehrke, S. Grigorieff, and J. É. Pin. "Duality and Equational Theory of Regular Languages". In: *Lect. Notes Comp. Sci. (ICALP)* 5125 (2008).

[5] S. Geman and M. Johnson. "Probabilistic Grammars and their Applications". In: *International Encyclopedia of the Social & Behavioral Sciences*. Ed. by N. J. Baltes and P. B. Smelser. Oxford: Pergamon, 2001. ISBN: 978-0-08-043076-8.

[6] J. Justesen and K. J. Larsen. "On probabilistic context-free grammars that achieve capacity". In: *Information and Control* 29.3 (1975). ISSN: 0019-9958.

[7] O. Kupferman, S. Safra, and M. Y. Vardi. "Relating word and tree automata". In: *Annals of Pure and Applied Logic* 138.1 - 3 (2006). New Zealand Institute of Mathematics and its Applications: Logic and Computation Programme. ISSN: 0168-0072.

[8] E. Leiss. "Succinct representation of regular languages by Boolean automata". In: *Theoretical Computer Science* 13.3 (1981). ISSN: 0304-3975.

[9] D. Perrin, J. É. Pin, and J. J. Rue. *Semigroups and Automata on Infinite Words*. 1995.

[10] J. É. Pin. *Mathematical foundations of automata theory*. 2014.

[11] J. É. Pin. *Mathematical grounds of automata theory*. 2009.

[12] N. Pippenger. "Regular Languages and Stone Duality". In: *Theory Comput. Syst.* 30.2 (1997).

[13] S. Salehi and M. Steinby. "Tree algebras and varieties of tree languages". In: *Theoretical Computer Science* 377.1 - 3 (2007). ISSN: 0304-3975.

[14] J. Skurczyński. "A characterisation of Büchi tree automata". In: *Information Processing Letters* 81.1 (2002). ISSN: 0020-0190.

[15] K. C. Vah and A. Salibra. "The connection of skew Boolean algebras and discriminator varieties to Church algebras". In: (2015). to appear.