Models of computation and finite automata

(MPRI course 2.16, 2021-2022)

Part II: Automata, Monoids and Logic - Problem Sheet 1

Note. The problems below vary greatly in difficulty: some are simple manipulations of the definitions or classical textbook results, others ask to fill in some details from proofs in the lectures, yet others require more serious thought, and some, marked with (*), are open problems and/or suggest research projects.

Exercise 1. Let Σ be a finite alphabet. Prove that the monoid Σ^* of finite Σ-words, has the following *universal property*: for any monoid M and any function $f\colon \Sigma \to M$, there exists a unique monoid homomorphism $\bar{f}\colon \Sigma^* \to M$ such that $\bar{f}(a) = f(a)$ for every $a \in \Sigma$.

Exercise 2. Let $h\colon \Sigma^* \to M$ be a homomorphism. Prove that, if $L = h^{-1}(P)$ for some $P \subseteq M$, then $h^{-1}(h(L)) = L$.

Exercise 3. Recall that a *non-deterministic finite automaton* (NFA) is a tuple $\mathcal{A} = (Q, \Sigma, \delta, I, F)$, where Q and Σ are finite sets, δ is a function from Σ to $\mathrm{Rel}(Q \times Q)$, and I and F are subsets of Q. Prove that, for any finite word $w \in \Sigma^*$, we have $\bar{\delta}(w) \cap (I \times F) \neq \emptyset$ if, and only if, there exists a finite path from a state $q \in I$ to a state $q' \in F$ in the labelled transition graph of $\mathcal{A}$.

Exercise 4. The aim of this exercise is to show that first-order variables and first-order quantifiers are superfluous in monadic second-order logic on finite words. Fix a finite alphabet Σ. Recall that, by definition, a *monadic second order formula* over Σ-words is built from atomic formulas $x \leq y$, $x \in X$, and $a(x)$ for each $a \in \Sigma$, by applying Boolean connectives, first-order quantifiers, and monadic second order quantifiers.
Define a *purely monadic second order formula* to be a formula built from atomic formulas $X \subseteq Y$, $X \leq Y$, $X \subseteq a$ for each $a \in \Sigma$, by applying Boolean connectives and monadic second order quantifiers. The semantics of the atomic formulas is defined by: $w, P, Q \models X \subseteq Y$ iff P is a subset of Q; $w, P, Q \models X \leq Y$ iff every position in P comes before every position in Q; $w, P \models X \subseteq a$ iff every position in P is labeled by an a in w.

a. Show that any purely monadic second order formula can be equivalently expressed using a monadic second order formlua.

b. Give a purely monadic second order formula $\mathrm{Sing}(X)$ which is true of the interpretation of X iff X is a singleton.

c. Describe an effective translation from monadic second order formulas to purely monadic second order formulas which preserves the satisfaction relation on finite words.

Exercise 5. Let $\mathcal{A} = (Q, \Sigma, \delta, I, F)$ be an NFA, with $Q = \{1, \ldots, n\}$. The aim of this exercise is to construct an MSO sentence ϕ such that a word $w \in \Sigma^+$ is accepted by $\mathcal{A}$ if, and only if, $w \models \phi$. The sentence ϕ will have the shape

$$\exists X_1 \ldots \exists X_n \psi,$$

where ψ is a first-order formula to be constructed below. The idea is that a successful run of the automaton $\mathcal{A}$ on a word w one-to-one corresponds to an interpretation of the second-order

variables $X_1, \ldots, X_n$ on w such that ψ is true. Here, the intended meaning of $p \in X_i$ will be that, on a successful run, the automaton is in state i immediately after reading the letter at position p in w.
For example, the formula

$$\psi_1 := \exists x\ (\forall y\ x \leq y) \wedge \bigvee_{a \in \Sigma} \bigvee_{j \in \delta(a)[I]} (a(x) \wedge x \in X_j)$$

expresses the fact that the run must begin in an initial state; here, $\delta(a)[I]$ denotes the set $\{j \in Q \colon \exists i \in I \text{ such that } (i,j) \in \delta(a)\}$.

a. Write a formula ψ_2 which expresses the fact that the run has to end in a final state.

b. Write a formula ψ_3 which expresses the fact that all the transitions in the run are legal according to the labeled transition graph of the automaton.

c. Write a formula ψ_4 which expresses the fact that the variables X_i are disjoint and cover the positions in the word.

d. Conclude that $\psi := \psi_1 \wedge \psi_2 \wedge \psi_3 \wedge \psi_4$ is a first-order formula as required.

e. In the above, we ignored the empty word. How would you account for it?

Exercise 6. This exercise fills in some steps of the proof that a language defined by an MSO formula can always be recognized by a finite monoid.

a. Let $\Sigma := \{a, b, c\}$. Compute a 5-element monoid M and a homomorphism from Σ^* to M which recognizes $\Sigma^* a \Sigma^* b \Sigma^*$.

b. Using the previous part, show that M recognizes the language defined by the atomic formula $X \leq Y$. *Hint.* This atomic formula is true for a word $w \in (2^{X,Y})^*$ iff there do not exist positions p, q in w with $w_p(X) = 1$, $w_q(Y) = 1$, and $q < p$.

c. Show that the collection of languages recognizable by a finite monoid is closed under binary intersections.

d. Let $n \geq 1$, M a finite monoid, and $h \colon (\Sigma \times 2^{\{X_1,\ldots,X_n\}})^* \to M$ a homomorphism. Define $H \colon (\Sigma \times 2^{\{X_1,\ldots,X_{n-1}\}})^* \to \mathcal{P}(M)$ by sending a letter (a,b) to $\{h(a, b \cdot 0), h(a, b \cdot 1)\}$. Prove that for any $w = (a_1, b_1) \ldots (a_m, b_m) \in (\Sigma \times 2^{\{X_1,\ldots,X_{n-1}\}})^*$, we have $H(w) = \{h(a_1 \ldots a_m, (b_1 \cdot c_1) \ldots (b_m \cdot c_m)) : c_1 \ldots c_m \in 2^m\}$.

Exercise 7. This exercise asks to show some basic properties of monoids and semigroups.

a. Prove that if a finite semigroup S is generated by a single element, then it is isomorphic to a cyclic semigroup $C_{c,d} := \{s, s^2, \ldots, s^{c+d-1}\}$, where $s^{c+d} = s^c$, and $c, d > 0$.

b. Prove that a semigroup is a group if and only if, for every element s, the maps $t \mapsto st$ and $t \mapsto ts$ are surjective.

c. Prove that a subgroup of a monoid does not need to be a submonoid.

d. Recall that e is called *idempotent* if $e^2 = e$. Prove that any element in a finite semigroup has a unique idempotent power.

e. Recall that a semigroup is called *aperiodic* if it has no subgroups of size > 1. Prove that a finite semigroup S is aperiodic if, and only if, there exists $n \geq 1$ such that $s^n = s^{n+1}$ for every element s of S.

Exercise 8. By a *marked starfree expression*, we mean one that is built from $\emptyset$ and $\{\epsilon\}$ by applying union, complement, and *marked concatenation*: if L, K are marked starfree expressions and $a \in \Sigma$, then LaK is also a marked starfree expression. Prove that marked starfree expressions describe the same class of languages as starfree expressions.

Exercise 9. Let $\phi(x_1, \ldots, x_n)$ be a first-order formula for finite Σ-words.

a. Show that there exists a formula $\phi^{<y}(x_1, \ldots, x_n, y)$ such that, for any finite word w and $p_1, \ldots, p_n, q \in |w|$,

$$w, p_1, \ldots, p_n, q \models \phi^{<y} \iff w(< q), p_1, \ldots, p_n \models \phi,$$

where $w(< q)$ denotes the length q prefix of w.

b. Formulate and prove analogous statements for the relativizations $\phi^{>y}$ to a suffix and $\phi^{(y,z)}$ to an infix.

Exercise 10. In this exercise, we write $\mathrm{FO}_{n,k}$ for the set of first-order formulas with free variables among $x_1, \ldots, x_n$ and of quantifier depth $\leq k$. Recall that, for any Σ-words u and v and markings $\bar{p} \in |u|^n$ and $\bar{q} \in |v|^n$, we write $u, \bar{p} \equiv_{n,k} v, \bar{q}$ iff the two marked words satisfy the same formulas in $\mathrm{FO}_{n,k}$. The aim of this exercise is to inductively construct, for every $n \geq 0$ and $k \geq 0$, a finite set $F_{n,k}$ of formulas in $\mathrm{FO}_{n,k}$ which *describe* the relation $\equiv_{n,k}$, in the following sense: for any marked Σ-word $(u, \bar{p})$, there is a unique $\phi \in F_{n,k}$ such that ϕ holds exactly in the marked Σ-words that are $\equiv_{n,k}$-equivalent to $(u, \bar{p})$.

a. Construct, for any $n \geq 0$, a set $F_{n,0}$ which describes $\equiv_{n,0}$.

b. Let $k \geq 0$ and assume by induction that, for any $n \geq 0$, a set $F_{n,k}$ describing $\equiv_{n,k}$ has been constructed. Let $n \geq 0$. For any subset S of $F_{n+1,k}$, consider the formula

$$\phi(S) := \bigwedge_{\psi \in S} \exists x_{n+1} \psi \wedge \forall x_{n+1} \bigvee_{\psi \in S} \psi.$$

Note that $\phi(S)$ is in $\mathrm{FO}_{n,k+1}$, and show that the set of formulas $\{\phi(S) : S \subseteq F_{n+1,k}\}$ describes $\equiv_{n,k+1}$.

c. Estimate the size of $F_{n,k}$ in terms of n and k. What is the complexity of computing the set $F_{n,k}$?

d. Let us say that a set $G_{n,k}$ of formulas *fully* describes $\equiv_{n,k}$ if it describes $\equiv_{n,k}$, and moreover, every formula in $G_{n,k}$ is satisfiable in some finite Σ-word. Modify your answers to the previous questions if the goal is to construct such a set $G_{n,k}$. For (c) this is probably (*) level.

e. Prove that any formula of $\mathrm{FO}_{n,k}$ is equivalent to a finite disjunction of formulas taken from $F_{n,k}$.

f. What changes if we want to fully describe MSO instead of FO?

Exercise 11. a. Prove that, for any $k \geq 1$ and any finite word $w \in \Sigma^*$, the ordered models associated to w^{2^k} and w^{2^k-1} satisfy the same formulas of $\mathrm{FO}_{0,k}$.

b. Show that the preceding claim is no longer true if one replaces FO by MSO.

Exercise 12. Show that the following theorem (a reformulation of Schützenberger's theorem) implies that any language recognized by a finite aperiodic semigroup is star-free.
Theorem. For any finite semigroup S, any finite subset $\Sigma \subseteq S$, and any $s \in S$, the set of words $u \in \Sigma^+$ that evaluate to s is star-free.

Exercise 13. Starting from the 5-element monoid M that you computed in Exercise 6, follow the proof of Schützenberger's theorem to see what star-free expression it yields for each element of the monoid.
Same question for a 6-element monoid M that recognizes the language $(ab)^* \subseteq \{a, b\}^*$.

Exercise 14. (*) Design and write a program that takes as input the multiplication table of a finite aperiodic semigroup S and a subset Σ of S, and produces a star-free expression for the set of words $u \in \Sigma^+$ that evaluate to s. Formally verify the correctness of your program.

Exercise 15. (*) Given a collection of groups $\mathbf{H}$, let $\bar{\mathbf{H}}$ denote the collection of finite monoids M such that all subgroups of M lie in $\mathbf{H}$. The collection $\bar{\mathbf{H}}$ generalizes the collection of aperiodic monoids, which is $\bar{\mathbf{H}}$ when $\mathbf{H}$ contains only the trivial group. What is the correct analogue of star-free expression and/or first-order logic for various choices of $\mathbf{H}$? Answers are known in the literature (only?) for the cases $\mathbf{H} = \{1\}$ (FO or star-free expressions), $\mathbf{H}$ is all finite groups (MSO or regular expressions), and $\mathbf{H}$ is all finite solvable groups (an extension of FO with "modular quantifiers", or modular counting expressions).

Exercise 16. (*) Generalize Schützenberger's theorem to other structures than finite words, including the ones naturally arising from the functorial point of view described in the first part of the semester, and/or those arising from the monadic point of view on automata theory. Answers, sometimes partial, are known in the literature for various kinds of infinite words, for some types of trees, for data words, for words over ordered alphabets, and more.