Problem Set 0

Instructions

Resources useful to solve the exercises in this problem set are the following:

Guille Godoy's video lectures

- Teoría de lenguajes (1)
- Teoría de lenguajes (2)
- Teoría de lenguajes (3)

Books

- (Sipser 2013, §0)
- (Cases and Màrquez 2003, §1)
- (Hopcroft, Motwani, and Ullman 2007, § 1)

Cases, Rafel, and Lluís Màrquez. 2003. *Llenguatges, Gramàtiques i Autòmats : Curs Bàsic*. 2a ed. Edicions UPC.

Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman. 2007. *Introduction to Automata Theory, Languages, and Computation*. 3rd edition. Pearson Addison Wesley.

Sipser, Michael. 2013. *Introduction to the Theory of Computation*. 3rd edition. Cengage Learning.

In the following exercises you can assume all well-known basic properties of union, intersection, and complement of sets. For example the distributivity of intersection over union, the distributivity of union over intersection, and de Morgan's laws, *i.e.* given sets A, B, C it holds that

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
- $A \cup (B \cap C) = (A \cap B) \cup (A \cup C)$,
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$,
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

Recall also that to show an equality between two sets A and B, it is enough to show that $A \subseteq B$ and $B \subseteq A$, that is for every $x \in A$ it holds that $x \in B$ and, viceversa, that for every $y \in B$ it holds that $y \in A$.

All exercises

Exercise 0.1 (From informal to formal descriptions)**.** Given an alphabet Σ, interesting languages $L \subseteq \Sigma^*$ can be often written using the set notation as $L = \{w \in \Sigma^* \mid P(w)\}$, where $P(w)$ is a predicate P defined on word w. For instance, the language L of all words over $\{a, b\}$ that contain the subword ab can be denoted as

$$L = \{w \in \{a, b\}^* \mid \exists x, y \in \{a, b\}^* \ w = xaby\}.$$

Write the formal languages over $\Sigma = \{a, b\}$ for the following informal descriptions of $P(w)$:

(a) To the right of every subword ab in w there is some subword ba.

(b) Word w contains both the subword ab and the subword ba.

(c) Between every two b's in w there is some a.

(d) Every occurrence of b in w is in an even position (the first symbol of a word is in position 1).

(e) Word w has some prefix with at least as many b's as a's.

(f) In every prefix of w, the number of b's is at least the number of a's.

(g) Word w has some even-length prefix with at least as many b's as a's.

(h) Every even-length prefix of w has at least as many b's as a's.

(i) Word w has a prefix and a suffix of the same nonzero length, equal to each other, and strictly less than the length of the word.

(j) Word w is a palindrome, that is, w reads the same forwards as backwards, e.g. *madam* or *rotator* in English.

> To define the property $P(w)$ formally, use universal and existential quantifiers ($\forall$, $\exists$), Boolean operators ($\vee$, $\wedge$, $\implies$, ...) and the notations about word lengths that we have introduced in class.

Exercise 0.2 (Concatenation – basic properties)**.** Justify your answers to the following questions. All words and languages are taken over a fixed alphabet.

(a) *Is the concatenation commutative?* Does it hold $xy = yx$ for all words x, y? Does it hold $AB = BA$ for all languages A, B? Do your previous answers depend on the size of the alphabet?

(b) *Is the concatenation associative?* Does it hold that $x(yz) = (xy)z$ for all words x, y, z? Does it hold that $A(BC) = (AB)C$ for all languages A, B, C? Do your previous answers depend on the size of the alphabet?

(c) *Does the cancellation law hold for concatenation of words?* That is, does $xy = xz$ imply $y = z$ for all words x, y, z?

(d) *Does the cancellation law hold for concatenation of languages?* That is, does $AB = AC$ imply $B = C$ for all languages A, B, C? What if we additionally impose that $A \neq \emptyset$, does it hold now?

(e) *Condition for a double equality.* Given languages A, B, C, D such that A, B are non-empty, $AB = CD$, and all words in A and C have the same length, does it hold that $A = C$ and $B = D$?

(f) *Does concatenation of languages distribute over union?* That is, for all languages A, B, C, does it hold that $(A \cup B)C = AC \cup BC$ and $A(B \cup C) = AB \cup AC$? What if instead of "$=$" we only ask for "$\subseteq$" or "$\supseteq$", does it hold now?

(g) *Does concatenation of languages distribute over intersection?* That is, does it hold that $(A \cap B)C = AC \cap BC$ and $A(B \cap C) = AB \cap AC$ for all languages A, B, C? What if instead of "$=$" we only ask for "$\subseteq$" or "$\supseteq$", does it hold now?

Exercise 0.3 (Kleene Star – basic properties)**.** Justify your answers to the following questions. Languages are taken over any fixed alphabet.

(a) *Does the Kleene star distribute over concatenation of languages?* That is, for every two languages L_1, L_2, does it hold that $L_1^* L_2^* = (L_1 L_2)^*$? What if instead of "$=$" we only asked for "$\subseteq$" or "$\supseteq$"? What if we impose $L_1 = L_2$, does it hold now? And what if, when $L_1 = L_2$, instead of "$=$" we only asked for "$\subseteq$" or "$\supseteq$"?

(b) *Do positive closures always induce nontrivial partitions?* That is, for every two languages L_1, L_2, if $L_1^+ \cup L_2^+ = \{a,b\}^+$ and $L_1^+ \cap L_2^+ = \emptyset$, then either $L_1 = \emptyset$ or $L_2 = \emptyset$.

(c) *Does the Kleene star distribute over union of languages?* That is, for every two languages L_1, L_2, does it hold that $L_1^* \cup L_2^* = (L_1 \cup L_2)^*$? What if instead of "$=$" we only asked for "$\subseteq$" or "$\supseteq$"?

(d) *Does the Kleene star distribute over intersection of languages?* That is, for every two languages L_1, L_2, does it hold that $L_1^* \cap L_2^* = (L_1 \cap L_2)^*$? What if instead of "$=$" we only asked for "$\subseteq$" or "$\supseteq$"?

(e) *Is the Kleene star monotone w.r.t. inclusion?* That is, for every two languages L_1, L_2, does $L_1 \subseteq L_2$ imply $L_1^* \subseteq L_2^*$? *Does the converse hold?* That is, for every two languages L_1, L_2, does $L_1^* \subseteq L_2^*$ imply $L_1 \subseteq L_2$? What if $L_1 = \{a,b\}$ and L_2 is a language over $\{a,b\}$, does the converse hold in this special context?

(f) *Chain of inclusions.* Does it hold for every language L that $\overline{L^*} \subseteq \overline{L} \subseteq \overline{L}^*$? *What if we reverse the inclusions?* Does $\overline{L^*} \supseteq \overline{L} \supseteq \overline{L}^*$ hold?

(g) *Sufficient condition for distinct Kleene stars.* Does $L_1 \cap L_2 = \emptyset$ for two languages $L_1, L_2 \neq \emptyset$ imply $L_1^* \neq L_2^*$?

(h) *Where is the positive closure squared?* Does $L^+ L^+ \subseteq L^+$ hold for every language L? Does the reverse inclusion $L^+ L^+ \supseteq L^+$ hold?

Exercise 0.4 (Characterizations)**.** Justify your answers to the following questions. Languages are taken over a fixed alphabet.

(a) *When is a language equal to its positive closure?* Does $L = L^+$ hold if and only if $L^2 \subseteq L$?

(b) *When is a language equal to its Kleene star?* Is $L^2 \subseteq L$ a necessary condition for the equality $L = L^*$? Is $\lambda \in L$ a necessary condition for $L = L^*$? Which logic combination ($\wedge$, $\vee$, ...) of the statements $L^2 \subseteq L$ and $\lambda \in L$ constitutes a necessary *and sufficient* condition for the equality $L = L^*$?

(c) *When is a language included in its square?* Is $\lambda \in L$ a sufficient condition for the inclusion $L \subseteq L^2$? Is $L = \emptyset$ a sufficent condition for $L \subseteq L^2$? Which logic combination ($\wedge$, $\vee$, ...) of the statements $L = \emptyset$ and $\lambda \in L$ constitutes a sufficient *and necessary* condition for the inclusion $L \subseteq L^2$?

(d) *When does a language equal its square?* Is $L = L^*$ a sufficient condition for the equality $L = L^2$? Is $L = \emptyset$ a sufficent condition for $L = L^2$? Which logic combination ($\wedge$, $\vee$, ...) of the statements $L = L^*$ and $L = \emptyset$ constitutes a sufficient *and necessary* condition for the equality $L = L^2$?

> Remember that if P and Q are two statements such that P implies Q, then P is said to be a *sufficient condition* for Q, while Q is said to be a *necessary condition* for P. In addition, P is called a *characterization* of Q when P is both a necessary condition and a sufficient condition for Q.

Exercise 0.5 (Reverse – basic properties)**.** Justify your answers to the following questions.

(a) *The reverse of the concatenation (1).* Show that for any two words x, y, $(xy)^R = y^R x^R$. *Does the analogue property hold for languages?* That is, given languages L_1, L_2, does it hold that $(L_1 L_2)^R = L_2^R L_1^R$?

(b) *The reverse of the concatenation (2).* Is it true that, given two languages L_1, L_2 such that $(L_1 L_2)^R = L_1^R L_2^R$, then necessarily $L_1 = L_2$?

(c) *Does the reverse distribute over union?* That is, given languages L_1, L_2, does it hold that $(L_1 \cup L_2)^R = L_1^R \cup L_2^R$?

(d) *Does the reverse distribute over intersection?* That is, given languages L_1, L_2, does it hold that $(L_1 \cap L_2)^R = L_1^R \cap L_2^R$?

(e) *Does taking the complement and the reverse of a language commute?* That is, it is true that $\overline{L}^R = \overline{L^R}$?

(f) *Does taking the Kleene star and the reverse of a language commute?* That is, it is true that $(L^*)^R = (L^R)^*$?

Exercise 0.6 (Homomorphisms I – basic properties)**.** Let $\sigma : \Sigma^* \to \Sigma^*$ be a function. Which of the following definitions of σ define a homomorphism? That is, which ones satisfy that for any words $x, y \in \Sigma^*$, $\sigma(xy) = \sigma(x)\sigma(y)$? Let $w = a_1 a_2 \cdots a_n \in \Sigma^*$, where $a_1, a_2, \dots, a_n \in \Sigma$.

(a) $\sigma(w) = a_1 a_1 a_2 a_2 \cdots a_n a_n$.

(b) $\sigma(w) = a_1 a_2 a_2 a_3 a_3 a_3 \cdots \overbrace{a_n \cdots a_n}^{n)}$.

(c) $\sigma(w) = ww$.

(d) $\sigma(w) = w$.

(e) $\sigma(w) = \lambda$.

(f) $\sigma(w) = a^{|w|}$, where $a \in \Sigma$.

(g) $\sigma(w) = w^R$.

(h) $\sigma(w) = \sigma_1(\sigma_2(w))$, where σ_1, σ_2 are homomorphisms.

Exercise 0.7 (Homomorphisms II – basic properties)**.** Given a morphism $\sigma : \Sigma^* \to \Sigma^*$ and languages $L, L_1, L_2 \subseteq \Sigma^*$, justify your answers to the following questions.

(a) *Does homomorphism on languages distribute over concatenation?* That is, does it hold that $\sigma(L_1 L_2) = \sigma(L_1)\sigma(L_2)$?

(b) *Does homomorphism and exponentiation of languages commute?* That is, does it hold that $\sigma(L^n) = \sigma(L)^n$ for any positive integer n?

(c) *Does homomorphism and union of languages commute?* That is, does it hold that $\sigma(L_1 \cup L_2) = \sigma(L_1) \cup \sigma(L_2)$?

(d) *Does homomorphism of languages and the Kleene star commute?* That is, does it hold that $\sigma(L^*) = \sigma(L)^*$?

(e) *Does homomorphism and reverse of languages commute?* That is, does it hold that $\sigma(L^R) = \sigma(L)^R$?

(f) *Does homomorphism and complementation of languages commute?* That is, does it hold that $\sigma(\overline{L}) = \overline{\sigma(L)}$?

(g) *Does the identity homomorphism on a language act as the identity on its words?* That is, does it hold that whenever $\sigma(L) = L$, then $\sigma(x) = x$ for all x in L?

Exercise 0.8 (On the size of languages)**.** Justify your answers to the following questions.

(a) Given two languages L_1, L_2, is it true that $|L_1| \cdot |L_2| = |L_1 \cdot L_2|$? What if $L_1 = L_2$?

(b) Given a homomorphism σ, is it true that if σ is injective then $|\sigma(L)| = |L|$?

(c) Given a language L, is it true that $|L^R| = |L|$?

(d) Given a language L and a positive integer n, is it true that $|L^n| = |L|^n$?

> Recall that a function f is *injective* if $f(x) = f(y)$ implies $x = y$.

Exercise 0.9 (Shifting a language)**.** Given a language L, we define the *shift* of L, denoted $S(L)$, as the language that contains the words obtained by applying a circular shift to each word in L in all possible ways; formally, $S(L) = \{vu \mid uv \in L\}$. Argue whether the following statements are true (with a justification) or false (with a counterexample) for any L.

(a) $S(L)^* \subseteq S(L^*)$.

(b) $S(L)^* \supseteq S(L^*)$.

(c) $\overline{S(L)} = S(\overline{L})$.

(d) $S(L^R) = S(L)^R$.

(e) $S(L_1 \cup L_2) = S(L_1) \cup S(L_2)$.

(f) $S(L_1 \cap L_2) = S(L_1) \cap S(L_2)$.

(g) $S(L_1 L_2) = S(L_1) S(L_2)$.

(h) $S(\sigma(L)) = \sigma(S(L))$, where σ is a homomorphism.

Exercise 0.10 (Simplification of languages)**.** Prove the following equalities between languages:

(a) $\{w \in \{a,b\}^* \mid aw = wb\} = \emptyset$.

(b) $\{w \in \{a,b\}^* \mid abw = wab\} = \{ab\}^*$.

> **Warning**
>
> $\{ab\}^*$ is not a typo, there is no "," between a and b.

(c) $\{xy \in \{a,b\}^* \mid |x|_a = |y|_a\} = \{w \in \{a,b\}^* \mid |w|_a \in 2\mathbb{N}\}$.

(d) $\{xy \in \{a,b\}^* \mid |x|_a = |y|_b\} = \{a,b\}^*$.

(e) $\{xy \in \{a,b\}^* \mid |x|_{aa} = |y|_b\} = \{a,b\}^*$.

(f) $\{w \in \{0,1\}^* \mid \texttt{value}_2(ww^R) \in 3\mathbb{N}\} = \{0,1\}^*$

> Note that the set notation $\{xy \in A \mid P(x,y)\}$, for a set A and a predicate P, is shorthand for $\{w \in A \mid \exists x, y\ w = xy \wedge P(x,y)\}$.

Exercise 0.11 ($L = \overline{\Sigma L}$)**.** Prove that, for every alphabet Σ, there is a unique language L such that $L = \overline{\Sigma L}$. What is this language?

Hint

Give an alternative expression for L. Does L contain the empty word λ? If a word in L is of the form aw, with $a \in \Sigma$, is w in L or in $\overline{L}$? Once obtained the new expression, rewrite it as a function of L and resolve the recurrence relation.

Exercise 0.12 (Arbitrary long walks in graphs)**.** Given a directed graph G (with loops) with n vertices, and two vertices s, t in it, show that if there exists a walk in G from s to t of length $> n$ then there are arbitrary long walks from s to t. In other words that there exists an $n_0 \in \mathbb{N}$ such that for every $m \geq n_0$ there is a walk in G from s to t of length $\geq m$.

> Recall that a *walk* from s to t in a graph G is a sequence of vertices $v_0, \ldots, v_\ell$ such that $v_0 = s$, $v_\ell = t$ and for each i, (v_i, v_{i+1}) is an edge in G. It is **not** required for the vertices or edges to be distict. The *length* of the walk is ℓ.

Exercise 0.13 (Easy induction)**.** Let $f : \mathbb{N} \to \mathbb{N}$ be a function such that $f(x+y) = f(x)+f(y)$ for any $x, y \in \mathbb{N}$. Prove that for any $x \in \mathbb{N}$, $f(x) = f(1) \cdot x$.

Exercise 0.14 (Divisibility in number set)**.** Prove that any set of $n+1$ positive integers less than or equal to $2n$, where $n \geq 1$, contains two distinct elements a and b such that a divides b.

Hint

A possibility is to use the pigeonhole principle: the pigeons are the $n+1$ numbers and the holes are the *odd* numbers between 1 and $2n$. A pigeon of the form $2^k d$ with d odd flies to hole d.
Another option is to consider an induction proof on n and distinguish two cases in the inductive step: at least one of the elements $2n+1$ and $2n+2$ is not in the set or both are.