Exercise 1.1.

Give a regular expression and an NFA for the language of all words over $\Sigma = \{a, b\}$ ...

(a) ...beginning and ending with $a$.

(b) ...such that the third letter from the right is a $b$.

(c) ...that can be obtained from $babba$ by deleting letters.

(d) ...with no occurrences of the subword $bba$.

(e) ...with at most one occurrence of the subword $bba$.

Solution. We write $\Sigma$ for $(a + b)$ and $\Sigma^*$ for $(a + b)^*$.

(a) $a + (a\Sigma^*a)$

One possible NFA for the language is the following. Note that every state of this NFA is initial and accepting. There are 7 states, labelled by 0, 1, 2, 3, 4, 5 and 6. From 0, upon reading a $b$, we can go to any state strictly bigger than 0; From 1, upon reading an $a$, we can go to any state strictly bigger than 1, and so on.
Exercise 1.2.

Let $A, B$ and $C$ be three languages.

(a) Prove that if $A \subseteq BC$ then $A^* \subseteq (B^* + C^*)^*$. Is the converse true?

(b) Prove that the languages of $((a + ba)^* b^*)$ and $(a + b)^*$ are the same.

Solution.

(a) Suppose $A \subseteq BC$. First, we show that $A^* \subseteq (BC)^*$. Indeed, if $w \in A^*$, then $w$ can be decomposed as $w_1w_2...w_n$ for some number $n$ such that each $w_i \in A$. Since $A \subseteq BC$, it follows that each $w_i \in BC$ and so $w \in (BC)^*$.

Now, we show that $(BC)^* \subseteq (B^* + C^*)^*$. If $w \in (BC)^*$ then $w$ can be decomposed as $w_1w_2...w_n$ for some number $n$ such that each $w_i \in BC$. Since each $w_i \in BC$, it follows that each $w_i$ can be further decomposed as $u_iv_i$ where $u_i \in B$ and $v_i \in C$. Hence $w = u_1v_1u_2v_2...u_nv_n$ and since each $u_i, v_i \in B + C \subseteq B^* + C^*$, it follows that $w \in (B^* + C^*)^*$.

(b) Let $U = (a + b), V = (a + ba)^*$ and $W = b^*$. We then have that $U \subseteq VW$ and so by the previous subpart, we have that $U^* \subseteq (V^* + W^*)^*$. Since $V^* = V$ and $W^* = W$, it follows that $(a + b)^* \subseteq ((a + ba)^* + b^*)^*$. Further, since $(a + b)^*$ is the set of all words over $\{a, b\}$, we have that $((a + ba)^* + b^*)^* \subseteq (a + b)^*$. The desired claim then follows.
Exercise 1.3.
Consider the language $L \subseteq \{a, b\}^*$ given by the regular expression $a^*b(ba)^*a$.
(a) Give an NFA that accepts $L$.
(b) Give a DFA that accepts $L$.

Solution.
(a) NFA accepting $L$

(b) DFA accepting $L$

Exercise 1.4.
Let $\Sigma = \{a, b\}$ and let $\Sigma^* = (a + b)^*$. Suppose $w = a_1a_2...a_n$ where each $a_i \in \Sigma$. Then the upward closure of a word $w$ is defined as the set
$$\uparrow w = \{u_1a_1u_2a_2...u_n a_n u_{n+1} : u_1, u_2, ..., u_{n+1} \in \Sigma^*\}$$

The upward closure of a language $L$ is defined as the set $\uparrow L = \cup_{w \in L} \uparrow w$.
(a) Give an algorithm that takes as input a regular expression $r$ and outputs a regular expression $\uparrow r$ such that $L(\uparrow r) = \uparrow (L(r))$.
(b) Give an algorithm that takes as input an NFA $A$ and outputs an NFA $B$ with exactly the same number of states as $A$ such that $L(B) = \uparrow L(A)$.

Solution.
(a) We define $\uparrow r$ by induction on the regular expression $r$:
- If $r = \emptyset$, then we set $\uparrow r = \emptyset$
- If $r = \epsilon$, then we set $\uparrow r = \Sigma^*$
- If $r = x$ for some $x \in \{a, b\}$, then we set $\uparrow r = \Sigma^* x \Sigma^*$
- If $r = r_1 + r_2$ for some $r_1$ and $r_2$, then we set $\uparrow r = (\uparrow r_1) + (\uparrow r_2)$
- If $r = r_1 r_2$ for some $r_1$ and $r_2$, then we set $\uparrow r = (\uparrow r_1)(\uparrow r_2)$
- If $r = (r_1)^*$ for some $r_1$, then we set $\uparrow r = \Sigma^*$. Note that if $r = (r_1)^*$ for some $r_1$, then $\epsilon \in L(r)$ and so $\uparrow L(r)$ must contain every word.
Let $A$ be an NFA recognizing a language $L$. We construct the NFA $B$ from $A$ as follows: Corresponding to every state $q$ of $A$ and every letter $x \in \{a, b\}$, we add a self-loop transition $(q, x, q)$. These new transitions will be referred to as \textit{special transitions}. We now claim that $\mathcal{L}(B) = \uparrow L$.

Suppose $w \in \uparrow L$. Hence, $w = u_1a_1u_2a_2...u_na_nu_{n+1}$ for some words $u_1, ..., u_{n+1}$ and letters $a_1, ..., a_n$ such that $u'_i := a_1a_2...a_n \in L$. Hence, there is an accepting run $\rho := q_0 \overset{a_1}{\rightarrow} q_1 \overset{a_2}{\rightarrow} q_2...q_{n-1} \overset{a_n}{\rightarrow} q_n$ of $A$ on the word $w'$. Now, notice that $q_0 \overset{a_1}{\rightarrow} q_0 \overset{a_2}{\rightarrow} q_1 \overset{a_3}{\rightarrow} q_1...q_{n-1} \overset{a_n}{\rightarrow} q_n \overset{u_{n+1}}{\rightarrow} q_n$ is an accepting run of $B$ on the word $w$. (Here $q_i \overset{u_{i+1}}{\rightarrow} q_i$ denotes that starting from the state $q_i$, there is a run on the word $u_{i+1}$ which ends at $q_i$). This implies that $w \in \mathcal{L}(B)$.

Suppose $\rho$ is an accepting run of $B$ on the word $w$. We now prove that $w \in \uparrow L$ by induction on the number of special transitions of $\rho$. If $\rho$ has no special transitions, then $\rho$ is also a run of $A$ on $w$ and so $w \in L \subseteq \uparrow L$. For the induction step, suppose $\rho$ has $k+1$ special transitions for some $k \geq 0$. Let $w = w_1w_2...w_n$ with each $w_i \in \Sigma$ and let $\rho = q_0 \overset{w_0}{\rightarrow} q_1 \overset{w_1}{\rightarrow} q_2...q_{n-1} \overset{w_n}{\rightarrow} q_n$. Let $q_i \overset{w_{i+1}}{\rightarrow} q_{i+1}$ be the first special transition along $\rho$. Since this is a special transition, it must be the case that $q_i = q_{i+1}$. Let $w'$ be the word obtained from $w$ by deleting the letter $w_{i+1}$ at the $(i+1)^{th}$ position and let $\rho'$ be the accepting run of $B$ on $w'$ obtained from $\rho$ by deleting the transition $q_i \overset{w_{i+1}}{\rightarrow} q_i$. Since $\rho'$ has only $k$ special transitions, by induction hypothesis, $w' \in \uparrow L$. Since $w$ can be obtained from $w'$ by adding a letter, it follows that $w \in \uparrow L$ as well, thereby finishing the proof.