CS3350

– Part 2 –

Regular languages
2.1 Regular expressions

Definition:
$E$ is a regular expression on alphabet $\Sigma$ iff

1. $E = \emptyset$, or
2. $E = a$ with $a \in \Sigma_{\varepsilon}$, or
3. $E = (E_1 \cup E_2)$ with $E_1$ and $E_2$ regular expressions, or
4. $E = (E_1 \cdot E_2)$ with $E_1$ and $E_2$ regular expressions, or
5. $E = (E_1)^*$ with $E_1$ a regular expression.

(inductive definition)
We often omit some of the parentheses. When parentheses are missing, the precedence order is: Kleene star, concatenation, union.

Example: \( a \cup ab^* \)

Other notations: +

1. \( E_1 + E_2 \equiv E_1 \cup E_2 \).
2. \( E^+ \equiv EE^* \).
Relation between a regular expression and the language it represents:
Let $E$ be a regular expression. We denote by $\mathcal{L}(E)$ the language represented by $E$, defined inductively as follows:

1. $\mathcal{L}(\emptyset) = \emptyset$
2. $\mathcal{L}(a) = \{a\}$ for $a \in \Sigma$
3. $\mathcal{L}(E_1 \cup E_2) = \mathcal{L}(E_1) \cup \mathcal{L}(E_2)$
4. $\mathcal{L}(E_1 \cdot E_2) = \mathcal{L}(E_1) \cdot \mathcal{L}(E_2)$
5. $\mathcal{L}(E_1^*) = \mathcal{L}(E_1)^*$. 
Examples on alphabet $\Sigma = \{a, b\}$

$$E = (a(a + b)^*) + ((a + b)^*a)$$

The language $\mathcal{L}(E)$ represented by $E$ is

$$= \{ w : w = a \circ x \text{ or } w = x \circ a \text{ with } x \in \Sigma^* \}.$$  

I.e. $\mathcal{L}(E)$ is the set of strings beginning or finishing by $a$.  

\[ E = ((a + b)(a + b))^* \]

\[ \mathcal{L}(E) = \{ w : |w| = 2k \text{ for } k \in \mathbb{N} \}. \]

\( E \) represents the set of strings of even length.
\[ E = (a + b)^* (aaa(a + b)^*) \]

\[ \mathcal{L}(E) = \{ w : w = x \circ aaa \circ y \text{ with } x, y \in \Sigma^* \} \]

\( E \) represents the language of strings containing three consecutive \( a \)'s.
Class exercises: Find regular expressions for the following languages Over \( \Sigma = \{a, b\} \):

1. The language of all strings where the length is a multiple of 3.
2. The language of all strings \( w \) where \( |w|_a \) is a multiple of 3.
3. The language of all strings that never have 2 consecutive a’s.
4. The language of all strings that never have 3 consecutive a’s.
**Definition**: Two regular expressions are equivalent if they represent the same language.

**Example**: \((a + b)^* \equiv (a*b*)^*\).

**Definition**: A language \(L\) is regular if there is a regular expression \(E\) such that \(\mathcal{L}(E) = L\).
2.2 An old problem

Alcuin (735–804)

The man, wolf, goat and cabbage
**States**: positions of man (H), wolf (L) goat (V) and cabbage (C) with respect to the river.

Example of state: “HV-LC”.

**Action**: the man crosses the river alone or accompanied (transition from a state to another).

Examples of an action:
“V” : the man crosses bringing the goat,
“L” : the man crosses bringing the wolf.
Another example of automaton:
And another...
2.3 Deterministic Finite Automata (DFA)

Formal definition: A deterministic finite automaton is a quintuple $(K, \Sigma, \delta, s, F)$, with

- $K$, finite set of states;
- $\Sigma$, alphabet (finite);
- $\delta: K \times \Sigma \rightarrow K$, transition function;
- $s \in K$, initial state;
- $F \subseteq K$, set of final states.
Example

Alphabet = $\Sigma = \{5c, 10c, 25c\}$.

Wanted: DFA that accepts the set of strings on alphabet $\Sigma$ such that the sum of the values of coins is less than $25c$.

$M = (K, \Sigma, \delta, q_0, F)$ where

$K = \{q_0, q_5, q_{10}, q_{15}, q_{20}, q_{25+}\}$

$F = \{q_0, q_5, q_{10}, q_{15}, q_{20}\}$

and $\delta$ is given by the following diagram:
Note: \(|L(M)| = ?\)

\(|L(M)| = 12 < \infty\).
**Definition:** Let $M = (K, \Sigma, \delta, s, F)$ be a DFA. A pair $[q, w]$ where $q \in K$ and $w \in \Sigma^*$ is a **configuration**.

If $a \in \Sigma$ and $\delta(q, a) = q'$ then

$$[q, a \circ w] \longrightarrow [q', w]$$

is the way to denote a simple transition from configuration $[q, a \circ w]$ to configuration $[q', w]$. 
For $x \in \Sigma^*$,

$$[q, x \circ w] \xrightarrow{n} [q', w]$$

denotes $n$ consecutive transitions.

Generally

$$[q, x \circ w] \xrightarrow{*} [q', w]$$

indicates $\exists n \geq 0$ such that $[q, x \circ w] \xrightarrow{n} [q', w]$.

(NB: $n = |x|$).
Let $M = (K, \Sigma, \delta, q_0, F)$ a DFA.

**Definition:** A state $q \in K$ is **reachable** if there exists $w \in \Sigma^*$ such that $[q_0, w] \xrightarrow{*}[q, \varepsilon]$.

**Definition:** Let $w \in \Sigma^*$. We say that $M$ **accepts** $w$ iff $\exists q \in F$ such that $[q_0, w] \xrightarrow{*}[q, \varepsilon]$.

**Definition:** Let $M$ be a DFA. $L(M) = \{w: M \text{ accepts } w\}$ is the **language accepted** by $M$. 
\[
\begin{align*}
[q_0, 10c5c10c] &\rightarrow [q_{10}, 5c10c] \rightarrow [q_{15}, 10c] \\
[q_0, 25c10c] &\rightarrow [q_{25+}, 10c] \rightarrow [q_{25+}, \varepsilon] \\
[q_0, 5c5c5c] &\rightarrow [q_5, 5c5c] \rightarrow [q_{10}, 5c] \rightarrow [q_{15}, \varepsilon]
\end{align*}
\]
Ex: even number of $a$’s on alphabet $\Sigma = \{a, b\}$.

Wanted: DFA $M$ such that $L(M) = \{w : |w|_a \text{ is even}\}$.

$M = (\{q_0, q_1\}, \Sigma, \delta, q_0, \{q_0\})$ where

$\delta$ is given formally by:

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_0$</td>
<td>$q_1$</td>
<td>$q_0$</td>
</tr>
<tr>
<td>$q_1$</td>
<td>$q_0$</td>
<td>$q_1$</td>
</tr>
</tbody>
</table>
With a diagram:

\[
\begin{array}{c}
\text{q}_0 \quad \text{a} \quad \text{q}_1 \\
\downarrow \quad \downarrow \quad \downarrow \\
\text{b} \quad \text{b} \\
\end{array}
\]

\[
\begin{array}{c}
\text{q}_0, aaaa \rightarrow [q_1, aaaa] \rightarrow [q_0, aa] \\
\text{q}_1, a \rightarrow [q_0, \varepsilon] \\
\text{q}_0, aba \rightarrow [q_1, ba] \rightarrow [q_1, a] \rightarrow [q_0, \varepsilon] \\
\text{q}_0, bab \rightarrow [q_0, ab] \rightarrow [q_1, b] \rightarrow [q_1, \varepsilon] \\
\end{array}
\]

Note: \(|L(M)| = ?\)

\(|L(M)| = \infty\).
Another example: $\Sigma = \{0, 1\}$

$M_1 = (\{q_1, q_2, q_3\}, \Sigma, \delta, q_1, \{q_2\})$

$\delta$ given by

$L(M_1) \supseteq \{101, 11001, 011, 1001, 101, 0^m1, \ldots, \}$

$w \in L(M_1) \Rightarrow w$ of the form $0^n1y$

$w \in L(M_1) \Rightarrow w$ of the form $0^n x 1(00)^k$
Would you put your hand in fire?

Then we need a proof!
To prove:

$w$ accepted by $M_1 \iff w = x \circ 1 \circ 0^m$ where $m$ even.

Preliminary facts:

1) $[q_2, 0^i] \xrightarrow{*}[q_2, \varepsilon] \implies i$ even  
   (induc. on “*”)

2) $i$ even $\implies [q_2, 0^i] \xrightarrow{*}[q_2, \varepsilon]$  
   (induc. on $i/2$)

3) $\forall y \in \Sigma^* \text{ et } \forall q$, $[q, y \circ 1] \xrightarrow{*}[q_2, \varepsilon]$  
   (inspection).
Proof of $\Rightarrow$ from the facts:

\begin{align*}
w \in L(M_1) &\Rightarrow w \text{ contains a } 1 \\
&\Rightarrow w \text{ of the form } y \circ 1 \circ 0^i \text{ with } i \geq 0 \\
&\Rightarrow [q_1, y10^i] \begin{array}{c} \ast \end{array} [q_2, 0^i] \begin{array}{c} \ast \end{array} [q_2, \varepsilon] \\
&\text{fact 3} \quad \text{w accepted} \\
&\Rightarrow i \text{ even by fact 1.}
\end{align*}
$w$ accepted by $M_1$ $\iff$ $w = x \circ 1 \circ 0^m$ where $m$ even

Proof of $\iff$ from the facts:

$w$ of the form $x \circ 1 \circ 0^m$ with $m$ even

$\Rightarrow [q_1, x10^m] \xrightarrow{*} [q_2, 0^m] \xrightarrow{*} [q_2, \varepsilon]$  

$\Rightarrow w \in L(M_1)$.  

\[ \textsf{fact 3} \quad \textsf{fact 2} \]
\[ \Sigma = \{a, b\} \]

\[ L(M_4) = ? \]

\[ L(M_4) = \Sigma \cup \{w: (w = a \circ x \circ a \text{ ou } w = b \circ x \circ b)\}. \]

\[ |L(M_4)| = \infty. \]
\[ \Sigma = \{0, 1\} \]
\[ L = \? \]

\[ L = \{w : w = x \circ 001 \circ y\} \]
\[ |L| = \infty. \]
2.4 Nondeterministic finite automaton

Definition:
A **nondeterministic finite automaton** (NFA) is a quintuple $(K, \Sigma, \Delta, s, F)$ where

- $K$ the finite set of states;
- $\Sigma$ is the alphabet (finite);
- $s$ is the initial state;
- $F \subseteq K$ is the set of final states;
- $\Delta \subseteq K \times (\Sigma \cup \{\varepsilon\}) \times K$ is the transition relation.
Nondeterministic finite automaton $= (K, \Sigma, \Delta, q_1, F)$ where $K = \{q_1, q_2, q_3, q_4\}$, $\Sigma = \{0, 1\}$, $F = \{q_4\}$

and $\Delta$ is formally given by

$\{(q_1, 0, q_1), (q_1, 1, q_1), (q_1, 1, q_2),$
$(q_2, 0, q_3), (q_2, \varepsilon, q_3),$
$(q_3, 1, q_4),$
$(q_4, 0, q_4), q_4, 1, q_4)\}$
**Definition**: Let $M = (K, \Sigma, \Delta, s, F)$ be a NFA. As in the case of the DFA, a **configuration** is a pair $[q, w]$ where $q \in K$ and $w \in \Sigma^*$. 
As for DFA’s, if $a \in \Sigma_\epsilon$ and $(q, a, q') \in \Delta$ then

$$[q, a \circ w] \xrightarrow{} [q', w]$$

is the way to denote going from configuration $[q, a \circ w]$ to configuration $[q', w]$.

NB:
- $[q, w] \xrightarrow{} [q', w]$ and $[q, w] \xrightarrow{} [q'', w]$ are possible for a NFA.
- $[q, w] \xrightarrow{} [q', w]$ is possible for a NFA.
As for DFAs, 
\[ [q, x \circ w] \xrightarrow{n} [q', w] \]
denotes \( n \) successive transition and 
\[ [q, x \circ w] \xrightarrow{*} [q', w] \]
indicates the existence of \( n \geq 0 \) such that 
\[ [q, x \circ w] \xrightarrow{n} [q', w] \].

(NB : \( n > |x| \) is possible for a NFA.)
Definition: Let $M = (K, \Sigma, \Delta, s, F)$ be a NFA and $w \in \Sigma^*$. As for a DFA, we say that $M$ accepts $w$ iff $[s, w] \xrightarrow{*} [q, \varepsilon]$ for some $q \in F$.

Definition: Let $M$ be a NFA with alphabet $\Sigma$. As for a DFA,

$$L(M) = \{w \in \Sigma^*: M \text{ accepts } w\}$$

is the language accepted by $M$. 
Ex : $\Sigma = \{0, 1\}$

$L(M) = \{w : w = x \circ 0101 \circ y\}$.

$M = (\{q_1, q_2, q_3, q_4, q_5\}, \{0, 1\}, \Delta, q_1, \{q_5\})$ where $\Delta$ is defined by the diagram:
Ex: $\Sigma = \{a, b\}$ and $L(M) = \{ w : |w|_a \equiv 0 \mod 2 \} \cup \{ w : |w|_b \equiv 0 \mod 4 \}$. 

\[
\begin{array}{c}
q_0 \xrightarrow{a} q_1 \xrightarrow{a} r_0 \xrightarrow{b} r_1 \xrightarrow{b} r_2 \xrightarrow{b} r_3 \xrightarrow{a} r_2 \xrightarrow{a} q_1 \xrightarrow{a} q_0
\end{array}
\]
Easy to construct a NFA for the union!
What if we wanted a DFA?

\[ \{ w : |w|_a \equiv 0 \mod 2 \} \cup \{ w : |w|_b \equiv 0 \mod 4 \}. \]

More complicated ... but possible.
How to go from a NFA to a DFA?

Is it always possible?

Let’s try on this example:
For example, consider say 1001, and see in which states the NFA can be after reading each symbol, one by one.

Color in blue the possible states of the NFA after reading $\varepsilon$, 1, 10, 100 and 1001.

After reading $\varepsilon$?
So, possible states after reading $\varepsilon$:

After reading 1?
So, possible states after reading 1:

After reading 10?

So, possible states after reading 10:

After reading 100?
So, possible states after reading 100:

After reading 1001?
So, possible states after reading 1001:

Is the string 1001 accepted by the NFA?

NO! Because the set of states we can reach after reading 1001 does not include a final state.
Is it possible in general to go from a NFA to a DFA? YES! Because...

- going from one line to the next in our calculations of which states of the NFA we can reach depends only of the symbol read, and

- the calculation from one line from the previous one is deterministic.

So, a DFA can simulate a NFA by “remembering” at all time the set of reachable states of the NFA!

We still need a more formal proof...
a DFA equivalent to the NFA from our example: