Logic, Descriptive Complexity and Theory of Databases

Lecture 8
Lecturer: Luc Segoufin
Scribe: Victor Marsault
November 22, 2010

Problem

INPUT : $L \in \text{Reg}$
OUTPUT : Is $L$ star-free ?

Definition

Finite State (Deterministic) Automaton

In the rest of this course notes, we denote by ‘automaton’ a FSA

$A =< \Sigma, Q, i, F, \delta >$

- $\Sigma$ : Some finite alphabet
- $Q$ : a finite set of state
- $i$ : a specific state ($\in Q$) called ‘initial’
- $F$ : $\subseteq Q$, a set of ‘final’ states
- $\delta$ : $Q \times \Sigma \rightarrow Q$

We extend $\delta$ to $Q \times (\Sigma)^+ \rightarrow Q$
st $\forall u, v \in (\Sigma)^+$, $\delta(q, uv) = \delta(\delta(q, u), v)$.

Definition

Periodicity

An automaton $A$ is periodic if.

$\exists (q_1, \cdots, q_k) \in Q^k$ with $k > 1$
$\exists u \in (\Sigma)^+$
$\{ \delta_A(q_1, u) = \delta_A(q_2, u), \cdots, \delta_A(q_{k-1}, u) = \delta_A(q_k, u), \delta_A(q_k, u) = q_1 \}$
$q_i \neq q_j$ if $i \neq j$

Furthermore, a regular language $L$ is periodic if the minimal deterministic automaton that recognizes $L$ is periodic.
**Definition**

Given a FSA $\mathcal{A}$ and $u \in (\Sigma_\mathcal{A})^*$:

Let $f_u : Q_\mathcal{A} \rightarrow Q_\mathcal{A}$ with $v \mapsto \delta_\mathcal{A}(q, u)$

**Problem**

PERIODICITY

INPUT : $\mathcal{A}$ a deterministic automaton

OUTPUT : Is it periodic ?

**Proposition**

PERIODICITY is decidable

First, it is easy to prove that:

$\mathcal{A}$ is periodic $\iff \exists u \in \Sigma_\mathcal{A}^+ (q_1, \cdots, q_k) \in Q_\mathcal{A}^k, k \geq 2$ such that

$f_u(q_1) = q_2, f_u(q_2) = q_3, \cdots, f_u(q_k) = q_1$.

Then it is enough to prove that one can compute $\{f_u\}_{u \in \Sigma_\mathcal{A}^+}$ from $\mathcal{A}$.

Let us prove that $\forall v \in \Sigma_\mathcal{A}^*, \exists u$, st $|u| \leq |Q|^{|Q|}, f_v = f_u$.

Let $v = v_1 \cdots v_n$ with $n > |Q|^{|Q|}$, and $v[k] = v_1 \cdots v_k$.

As $n \geq |Q|^{|Q|}$, there exists $i < j$ such that $f_v[i] = f_v[j]$ with $v = \underbrace{v_1 \cdots v_i \cdots v_j \cdots v_n}_{v[i]}$.

Let $g(v) = v_1 \cdots v_i v_j+1 \cdots v_n$, then $f_v = f_{g(v)}$, and $|g(v)| < |v|$.

One can iterate this process until the size is $\leq |Q|^{|Q|}$.

This shows that it is enough to consider $u$ of length bounded by $|Q|^{|Q|}$ in order to obtain all possible $f_u$.

**Theorem**

$L$ definable in $FO(<) \iff L$ is aperiodic (not periodic)

Case $\Rightarrow$ :

Towards a contradiction, let us assume $L$ periodic (of minimal FSA $\mathcal{A}$ and period $(u, (q_1, q_2, \cdots, q_k))$, and $L$ definable in $FO(<)$. Let $\varphi$ be the $FO(<)$ formula defining $L$, and $l = qr(\varphi)$.

Let us denote by $q \sim_A p$ if $\forall v \ [\delta_\mathcal{A}(q, v) \in F_\mathcal{A} \iff \delta_\mathcal{A}(p, v) \in F_\mathcal{A}]$.

Note that since $\mathcal{A}$ is minimal, there does NOT exists distinct states $p, q$ with $q \sim_A p$.

Let $v_1$ be any word st $\delta_\mathcal{A}(i_A, v_1) = q_1$ (it exists because all states of $\mathcal{A}$ are reachable by minimality of $\mathcal{A}$).

Let $v$ be such that $\delta_\mathcal{A}(q_1, v) \in F_\mathcal{A}$. Hence $v_1 v \in L$. By periodicity we actually have $\forall n \ v_1 u^{kn} v \in L$.

Let $n$ st $kn > 2^l$, then (since $u^{2^l} \equiv_l u^{(2^l+1)}$, $v_1 u^{kn} v \equiv_l v_1 u^{kn+1} v$). As we have $v_1 u^{kn} v \in L$ we therefore conclude that $v_1 u^{kn+1} v \in L$ and $\delta_\mathcal{A}(i_A, v_1 u^{kn+1} v) \in F_\mathcal{A}$.
Since $\delta_A(i_A, v_1 u^k n) = q_1$, $\delta_A(i_A, v_1 u^{k+1}) = q_2$
And since $\delta_A(i_A, v_1 u^{k+1} v) \in F_A$, $\delta_A(q_2, v) \in F_A$.
We just proved that $\forall v$ such that $\delta_A(q_1, v) \in F_A$, $\delta_A(q_2, v) \in F_A$.
By symmetry we have: $\forall i \leq k$, $\forall v$ such that $\delta_A(q_i, v) \in F_A$
$\delta_A(q_{i+1} \mod (k+1)), v) \in F_A$.
Which is enough to deduce $q_1 \sim_A q_2 \sim_A \cdots q_k$.
$\rightarrow$ Contradiction with the minimality of $A$.

Case $\Leftarrow$: We assume that $L$ is aperiodic.
Let $m = |\Sigma_A|$ and $n = |Q|$.
Sketch of the proof by induction on $(n, m)$ (with lexicographic order)
Subcase $\Leftarrow 1$: $\forall a \in \Sigma_A, f_a$ is onto.
Then $f_a$ is a permutation on $Q$. If it contains a non trivial cycle, then this
cycle contradicts aperiodicity (a with the cycle would form a period).
Hence $f_a$ is the identity on $Q$.
Since $\forall a \in \Sigma_A, f_a = Id$, $L = \Sigma_A^*$ or $L = \emptyset$ and both are definable in
$FO(<)$
Subcase $\Leftarrow 2$: $\exists a \in \Sigma_A$ st $f_a(Q) \subseteq Q$.
Let $B = \{ua \mid u \in (\Sigma_A - \{a\})^*\}$
and $\Gamma = \{f_a \mid x \in B\}$
and $L' = \{w \in \Gamma^* \mid wx \in L \text{ and } x \in (\Sigma_A - \{a\})^*\}$
Claim: $L'$ is aperiodic and its min FSA uses only the states in $f_a(Q)$.
(Only treating the case st $i_A \in f_a(Q_A)$ and $F_A \cap f_a(Q_A) \neq \emptyset$)
Let $A$ recognizing $L$.
Let’s build $A'$ recognizing $L'$ with:
$\begin{align*}
\Sigma_{A'} &= \Gamma \\
Q_{A'} &= f_a(Q) \\
i_{A'} &= i_A \\
F_{A'} &= F_A \cap f_a(Q_A) \\
\delta_{A'}(a, f) &\rightarrow f(q)
\end{align*}$
If $\gamma$ is a period-word of $A'$, then any $u$ st $f_u = \gamma$ is a period-word of $A$.
(Note that minimizing $A'$ cannot create any period). Let $\psi$ be the
$FO(<)$ formula for $L'$ (exists by induction hypothesis as $n$ decreased).
Consider now: $\forall \gamma \in Q^Q$, $q \in Q$, let us define $L^\gamma_q = \{v \in (\Sigma_A - \{a\})^* \mid f_{iA}(q) = \gamma(q)\}$
Claim: $\forall \gamma, q L^\gamma_q$ is aperiodic using $|Q|$ states.
Let’s build $A''$ with:
$\begin{align*}
\Sigma_{A''} &= \Sigma_A - \{a\} \\
Q_{A''} &= Q_A \\
i_{A''} &= q \\
F_{A''} &= \{p \in Q_A \mid \delta_A(p, a) = \gamma(q)\} \\
\delta_{A''}(a, \delta) &= (\delta_A)(Q \times (\Sigma_A - \{a\})
\end{align*}$
If $L^\gamma_q$ is periodic, then $A''$ is periodic (minimizing cannot create peri-
ods), then $A$ is periodic. Let $\phi^\gamma_q$ its $FO(<)$ formula (exists by induction
hypothesis, $n$ is the same and $m$ decreased as letter $a$ disappeared).
Let $L_\gamma = \bigcap_{q \in Q} L_\gamma^q = \{ v \in (\Sigma A - \{a\})^* \mid \forall q f_{va}(q) = \gamma(q) \}$ and

$\varphi_\gamma = \bigwedge_{q \in Q, A} \varphi_\gamma^q$.

From $\psi$ and $(\varphi_\gamma)_{\gamma \in \Gamma}$ computed above, we construct the desired $FO(<)$-formula defining $L$.

In $\psi$ we replace (careful, below we don’t consider two special cases: the initial and final segments of the word before the first $a$ and after the last $a$. These special cases should be treated using a disjunction)

- each quantification $\exists x\varphi$ by $\exists x, a(x) \land \varphi$
- each quantification $\forall x\varphi$ by $\forall x, a(x) \Rightarrow \varphi$
- each atom $P_\gamma(x)$ by $\exists y a(y) \land (y < x) \land (\forall z (y < z < x) \Rightarrow \neg a(z)) \land \varphi_\gamma^{[y,x]}$ where $\varphi_\gamma^{[y,x]}$ is $\varphi_\gamma$ in which one restricts all quantification to the interval between $y$ and $x$ (ie replacing every $\exists z \varphi'$ by $\exists z (y < z < x) \land \varphi'$ and every $\forall z \varphi'$ by $\forall z (y < z < x) \Rightarrow \varphi'$.