Logic, Descriptive Complexity and Theory of Databases

Lecture : 7
Lecturer : Luc Segoufin
Scribe : Amit Kumar Dhar

Monday November 15th

1 Recap

From the previous classes we have the following results:

FO $\subseteq$ LogSpace, follows 0-1 law, has locality and E-F Games
ESO = NP
LFP = IFP $\subseteq$ PFP.
LFP, IFP $\subseteq$ PTime.
PFP $\subseteq$ PSpace.
[Immerman-Vardi Theorem] In case of ordered Structure,
LFP = IFP = PTime
PFP = PSpace.
[Abiteboul-Vianu Theorem] LFP = PFP iff PTime = PSpace.

2 Monadic Second Order Logic (MSO)

As defined in the last class, MSO extends FO with quantification over sets. The general syntax is as follows : $\exists$S; $\forall$S; $\exists$x; $\forall$y; S(x); x $\in$ S; $\wedge$; $\vee$; $\neg$; atoms.
We have the following observations about MSO :

- Not well behaved to fit the above picture.
- MSO does not have 0-1 laws.
- MSO $\subseteq$ PSpace
  In particular MSO contains a complete problem for any level of polynomial hierarchy, and thus MSO $\subseteq$ PH

2.1 EF Games For MSO

Notations: Monadic EF Games represented by $\text{MEF}_n(\mathcal{A}, \mathcal{B})$ consists of

- Boards : $\mathcal{A}, \mathcal{B}$
- 2 Players.
- n moves
  where move i can be one the following type:
  - FO move:
    * Sp place pebble i on an element of $\mathcal{A}$.
    * D responds with placing pebble i on an element of $\mathcal{B}$.
    * and vice versa interchanging the role of $\mathcal{A}$ and $\mathcal{B}$.
  - MSO move:
    * Sp chooses a set S_i of elements of $\mathcal{A}$.
    * D responds with choosing a set S_i of element of $\mathcal{B}$.
    * and vice versa interchanging the role of $\mathcal{A}$ and $\mathcal{B}$.
- Winning condition
  a tuple $\overline{x}$ was selected in $\mathcal{A}$ while playing the game.
  a tuple $\overline{y}$ was selected in $\mathcal{B}$ while playing the game.
  D wins if there is a partial isomorphism between the structures expanded with the sets chosen.
  $(\mathcal{A}, \overline{S_1})/\overline{x} \to (\mathcal{B}, \overline{S_2})/\overline{y}$ i.e. $\mathcal{A} \models R(\overline{x})$ iff $\mathcal{B} \models R(\overline{y})$ and $x \in S_i$ iff $y \in S_i$.

We say, $\mathcal{A} \equiv_n^{MSO} \mathcal{B}$ if D has a winning strategy in $MEF_n(\mathcal{A}, \mathcal{B})$.

Theorem 1. *$\mathcal{A} \equiv_n^{MSO} \mathcal{B}$ iff $\forall\ \varphi \in MSO$, $qr(\varphi) \leq n : \mathcal{A} \models \varphi$ iff $\mathcal{B} \models \varphi$.*

Proof. The "if" direction of the proof is similar to the proof as in the case of EF games for FO. ☐

Let MSO_n be defined as $MSO_n = \{\varphi \in MSO | qr(\varphi) \leq n\}$.
Properties :

1. $\forall n$ MSO_n expresses only a finite number of **properties**.
   Note that there can be infinitely many formulaes.
2. $\forall n$ the equivalence relation $\equiv_n^{MSO}$ has finitely many equivalence classes.
3. $\forall n$ each class of $\equiv_n^{MSO}$ is expressible in MSO_n.
   where expressible means :
   - For each class c of $\equiv_n^{MSO}$ $\exists \psi \in MSO_n$ such that $\forall \mathcal{A} : \mathcal{A} \models \psi$ iff $\mathcal{A} \in C$

- $\forall \mathcal{A}\ \exists \psi \in MSO_n\ \forall \mathcal{B}\ \mathcal{A} \equiv_n^{MSO} \mathcal{B}$ iff $\mathcal{B} \models \psi$

Proof of Property 2. $\mathcal{A} \not\equiv_n^{MSO} \mathcal{B}$ iff $\exists \varphi \in MSO_n : \mathcal{A} \models \varphi \wedge \mathcal{B} \models \neg\varphi$ (By Theorem 1)
But there are finitely many possibilities for φ by Property 1.
Thus, together they imply that there are finitely many equivalence classes of $\equiv_n^{MSO}$. $\square$

Proof of Property 3. We, fix a model $\mathcal{A}$.
Let $X = \{\varphi \in MSO_n, \mathcal{A} \models \varphi_n\}$
By Property 1, X is finite.
We construct a formula ψ as :

$$\psi = \bigwedge_{\varphi \in X} \varphi$$

Clearly, $\psi \in MSO_n$ and for any other model $\mathcal{B}$, we have:
The "if" part of the property follows directly from Theorem 1.
On the other hand, if $\mathcal{B} \models \psi$ then $\mathcal{B}$ satisfies all the formulas of MSO_n that are satisfied by $\mathcal{A}$. Thus, $\mathcal{A} \equiv_n^{MSO} \mathcal{B}$. $\square$

Proof Sketch of Property 1. It can be proved induction on the quantifier ranks n of formulas.
For n=1,

we cannot start with a second order quantification, as we cannot use free variables to give a meaningful formula.

We can start with first order quantification followed by a property expressible without quantifiers, which is finite.

By inducting on n we can prove Property 1. $\square$

2.2 MSO on Words

In this case we deal with the following schema: $\sigma : \{<, P_a, P_b\}$ with $\Sigma = \{a, b\}$.
We use the following notational shortcut :

- $+1$ for denoting successor in linear order
- min for denoting the first position of the word.
- max for denoting the last position of the word.
- $a(x)$ for denoting the predicate $P_a(x)$.

A sentence φ, defines a language $L_\varphi = \{w \in \Sigma^* | w \models \varphi\}$

Theorem 2. *A language is definable in MSO iff it is REGULAR.*
(i.e. in terms of language equivalence MSO $\equiv$ REG).

Proof. :
($\Leftarrow$) : Let A $= < \Sigma, Q, q_0, F, \delta >$ be a finite state automaton.
where $Q = \{q_0, q_1, \cdots, q_k\}$
We construct the following MSO formula :
$\exists S_0 \exists S_1 \exists S_2 \cdots \exists S_k \psi$
where ψ says:

- S_i partitions the domain.
- $min \in S_0$
- $\forall xy, y = x + 1 \rightarrow \bigvee_{(q,a,q') \in \delta} a(x) \wedge S_q(x) \wedge S_{q'}(y)$
- $\bigvee_{(a,p) \exists q \in F, (p,a,q) \in \delta} max \in S_p \wedge a(max)$

Each S_i in the above formula is for each of the states of the automata such that $S_i = \{x : A \text{ is in state } q_i \text{ at } x\}$. And thus ensuring that the S_i partitions the domain gives us that the automata is only in one state at any point of time.
The formula ψ simulates the run of the automata on a given word and is satisfied only if the automata stops in some final state.
With the existential quantifiers on sets, it guesses a run of the following form on the word $q_0 a_0 q_1 a_1 \cdots a_n q_n \in F$. It then verifies the following :

- The run starts from the initial state of the automaton.
- It follows the transition rule of the automaton.
- It ends at a final state of the automaton.

By construction, the formula is satisfied if there exists an accepting run of the automaton.
($\Rightarrow$)
For proving the other direction we need the following lemma:

Lemma 1 (Composition Lemma). *If $u \equiv_n^{MSO} v$ and $u' \equiv_n^{MSO} v'$ then $u.v \equiv_n^{MSO} u'v'$ and $\langle u.v, |u| \rangle \equiv_n^{MSO} \langle u'.v', |u'| \rangle$*

Proof. Assume that Duplicator D has a winning strategy for $MEF_n(u, u')$ and $MEF_n(v, v')$ We will give a winning strategy for duplicator D in $MEF_n(u.v, u'.v')$. We will assume that the Spoiler Sp plays a move in $u.v$, since the other case is symmetric.

- FO Moves
  Say Sp selects a position x.
  if $x \leq |u|$ then D gets a $y \leq |u'|$ by using its strategy in $MEF_n(u, u')$.
  Similarly for $x > |u|$, D gets a $y > |u'|$ by using its strategy in $MEF_n(v, v')$.
  Within u or v the order is preserved, and it is preserved by the definition of concatenation in case x and y are from different parts.

- MSO Moves
  Say Sp selects a set S.
  S can be split as disjoint union of two sets as $S = S_1 \biguplus S_2$ where $S_1 = \{x \in S \mid x \leq |u|\}$ and $S_2 = \{x \in S \mid x > |u|\}$.
  D gets $T_1[T_2]$ for $S_1[S_2]$ using its strategy in $MEF_n(u, u')[MEF_n(v, v')]$
  We take the set chosen by D as $T = T_1 \biguplus T_2$

Hence, we have $\langle u.v, |u|\rangle \equiv_n^{MSO} \langle u'.v', |u'|\rangle$. □

Let $\varphi \in MSO$, and $n = qr(\varphi)$.
We construct an automaton A recognising the language L_φ as follows:
$A = (\Sigma, Q, q_0, F, \delta)$ where,
$Q = \{\tau : \tau \text{ is an equivalence class of } \equiv_n^{MSO}\}$
q_0 = equivalence class corresponding to ε.
$F = \{\tau : \mathcal{A} \in \tau, \mathcal{A} \models \varphi\}$
$\delta(\tau, a) = \tau' = [u.a]$ where $u \in \tau$ and $[u.a]$ denotes the $\equiv_n^{MSO}$-class of ua.
To show that A recognises L_φ, we take a run ρ of automaton A on w. By induction on x, we will prove that, A is in state τ at x iff $w[1 \cdots x] \in \tau$ (where $w[1 \cdots x]$ is the prefix of w till x)
Obviously, by construction it is true for q_0.
The induction step follows from the composition lemma. □

Corollary :

- REGULAR languages are closed under $\wedge$, $\vee$, $\neg$.
- Automata are determinizable.
- MSO=EMSO (on words)

Application :

Theorem 3. *Hamiltonicity $\notin$ MSO.*

Proof. We take a complete bipartite graph $K_{m,n}$. We know that $K_{m,n}$ has a hamiltonian cycle iff $m = n$.
Assume that there exists a formula $\varphi \in MSO$ which defines hamiltonicity. We take $\Sigma = \{a, b\}$ and define $\psi(x, y) \in FO(words) : ((P_a(x) \wedge P_b(y)) \vee (P_b(x) \wedge P_a(y)))$
Now, we construct another formula φ' which is obtained by replacing each occurrence of $E(x, y)$ by $\psi(x, y)$.
φ' is a MSO formula on words such that $w \models \varphi'$ iff $\#a = \#b$. But this language is not regular. Thus, by theorem 2, our assumption was wrong.
Hence there cannot be any such φ to express hamiltonicity. □

3 FO on Words

Here we look at first order logic with order relation, i.e. $FO(<)$.

3.1 Star Free Expressions

Star-free expressions are constructed from the letters of the alphabets, union, intersection, concatenation, complement and the empty set. Languages for which there exists a star-free expression are called star-free languages.

Example : $(ab)^* = a\Sigma^* \cap \Sigma^* b \cap \neg(\Sigma^* aa\Sigma^*) \cap \neg(\Sigma^* bb\Sigma^*)$ where $\Sigma^* = \emptyset^c$

Theorem 4. *A language is definable in FO iff it is star-free.*
(i.e. in terms of language equivalence FO $\equiv$ star-free languages).

Proof. :
($\Leftarrow$) We can prove this using induction on the expression.
The base step is obvious, and so is the induction step for union, intersection, complement as FO formulas are closed under these operations. The induction step for concatenation is as follows:
Let us consider two expressions e and e'. By induction hypotheses, we have FO formulas φ_e and $\varphi_{e'}$ respectively.
The FO formula for $e.e'$ is given by $\exists x \varphi_e^{\leq x} \wedge \varphi_{e'}^{>x}$.
where, $\varphi_e^{\leq x}$ is obtained from φ_e by replacing each $\exists z$ with $\exists z(z \leq x)$
and $\varphi_{e'}^{>x}$ is obtained from $\varphi_{e'}$ by replacing each $\exists z$ with $\exists z(z > x)$
Thus, we have a FO formula for every star-free expression.

($\Rightarrow$) We can prove this by induction on the quantifier rank of the formula.
All cases are easy except the case where the formula is of the type $\psi(x) = \exists x \varphi(x)$
Before that we will state the corresponding properties and the composition lemma for FO.

Let FO_n be defined as $FO_n = \{\varphi \in FO | qr(\varphi) \leq n\}$. And $\equiv_n^{FO}$ defined similarly as in the case of MSO, only on EF games of FO.
Properties :

1. $\forall n$ FO_n expresses only a finite number of **properties**.
2. $\forall n$ the equivalence relation $\equiv_n^{FO}$ has finitely many equivalence classes.
3. $\forall n$ each class of $\equiv_n^{FO}$ is expressible in FO_n.

Lemma 2 (**Composition Lemma for FO**). *If $u \equiv_n^{FO} v$ and $u' \equiv_n^{FO} v'$ then $u.v \equiv_n^{FO} u'v'$ and $\langle u.v, |u|\rangle \equiv_n^{FO} \langle u'.v', |u'|\rangle$*

Proof. It can be proved in the same way as in the case of MSO, by giving a similar strategy for the duplicator for EF games of FO by excluding the MSO moves. $\square$

Now, we construct the following set
$S_\psi = \{(\tau_1, \tau_2) | \exists w_1, w_2, w_1 \in \tau_1, w_2 \in \tau_2, w_1.w_2 \models \varphi(|w_1|)\}$
where τ_1, τ_2 are equivalence classes of $\equiv_n^{FO}$ and n is the quantifier rank of φ

S_ψ consists of the pair of equivalence classes of the different ways w can be partioned into two parts. The position of split is considered as x.

Claim 1. $L_\psi = \bigcup_{(\tau_1,\tau_2)\in S_\psi} L_{\tau_1}.L_{\tau_2}$
where L_α is the language of the FO_n formula expressing the class α obtained by Property 3 above.

Proof. :
($\Rightarrow$) Assume $w \models \psi$
Therefore, $\exists w \models \varphi(x)$. Thus, we take $w_1 = w[1\cdots x]$ and $w_2 = w[x+1\cdots max]$ and τ_1, τ_2 to be the equivalence class of w_1, w_2 respectively.
So, $(\tau_1,\tau_2) \in S_\psi$, $w_1 \in L_{\tau_1}$, $w_2 \in L_{\tau_2}$ implies that $w \in \bigcup_{(\tau_1,\tau_2)\in S_\psi} L_{\tau_1}.L_{\tau_2}$

($\Leftarrow$) Assume $w \in \bigcup_{(\tau_1,\tau_2)\in S_\psi} L_{\tau_1}.L_{\tau_2}$
Thus, w can be split into v_1 and v_2 such that $w = v_1.v_2$ and $v_1 \in L_{\tau_1}$ and $v_2 \in L_{\tau_2}$ for some $(\tau_1,\tau_2) \in S_\psi$.
But, $(\tau_1,\tau_2) \in S_\psi$ implies that by definition there exists w_1, w_2 such that, $w_1 \in \tau_1, w_2 \in \tau_2, w_1.w_2 \models \varphi(|w_1|)$
Again, by induction hypothesis, $v_1 \equiv_n^{FO} w_1$ and $v_2 \equiv_n^{FO} w_2$ as they belong to the same equivalence classes and hence by composition lemma, $\langle v_1.v_2, |v_1|\rangle \equiv_n^{FO} \langle w_1.w_2, |w_1|\rangle$
By the theorem on EF-games, this means that $\langle v_1.v_2, |v_1|\rangle$ and $\langle w_1.w_2, |w_1|\rangle$ satisfy the same formulas of quantifier rank less than n.
We know from the definition of S_ψ that $w_1.w_2 \models \varphi(|w_1|)$ and that φ has quantifier less than n. Hence from above, we have $v_1.v_2 \models \varphi(|v_1|)$
Thus, $w \models \exists x\varphi(x)$. □

Clearly, from the above claim and by induction hypothesis we have star-free expressions for L_{τ_1}, L_{τ_2} and using the fact that star-free expressions are closed under $\cup$ and concatenation, gives us the proof that L_ψ is star-free. □

Corollary :

- $(ab)^* \in FO$ as it is star-free as seen in the example.
- $(aa)^*$ is not star-free as it is not in FO.