Questions and Answers that arose While Teaching Formal Language Theory

Richard Beigel *

Temple University

William Gasarch †

Univ. of MD at College Park

1 Introduction

When teaching formal language theory many questions come up. In this short note we ask and answer some of them, and leave some open problems.

2 A Proof that REG is Closed Under Concatentation and Star that only Uses DFA's

In class I often say that the best way to prove that REG is closed under concatentation (star) is to use the Regular Expression Formulation or use NFA's. But what if you only knew about DFA's? Can one show closure under concatenation? Star? Yes!

Theorem 2.1 *If L_1 is DFA-regular and L_2 is DFA-regular then L_1L_2 is DFA-regular.*

Proof:

Let L_1 be regular via $M_1 = (Q_1, \Sigma, \delta_1, s_1, F_1)$. Let L_2 be regular via $M_2 = (Q_2, \Sigma, \delta_2, s_2, F_2)$. We construct a DFA for L_1L_2. The idea is that we will keep track of what state we would be if we

*Dept. of Computer and Information Sciences, Temple University, 1805 N. Broad St, Philadelphia, PA 19122. professorb@gmail.com

†University of Maryland, College Park, MD 20742, gasarch@cs.umd.edu

were in M_1 and also what SET of states we could be in M_2 if we assumed M_2 began working on the string after some final state in M_1.

$$(Q_1 \times 2^{Q_2}, \Sigma, \delta, (s_1, \emptyset), F)$$

where δ is defined as

$\delta((q, T, \sigma) = (\delta_1(q), \delta_2(T, \sigma))$ if $q \notin F_1$

$\delta((q, T, \sigma) = (\delta_1(q), \delta_2(T, \sigma) \cup \delta_2(s_2, \sigma))$ if $q \in F_1$

and

$$F = \{(q, T) \in Q_1 \times 2^{Q_2} \mid T \cap F_2 \neq \emptyset\}$$

▮

This proof results in the number of states being exponetial in the number of states in M_2. Is this needed? Yes: Look at

$L_1 = \{a, b\}^*$. The min DFA for L_1 has 2 states.

$L_2 = \{\{a, b\}^n\}^*$. The min DFA for this has n states.

It is known that $L_1 L_2$. requires 2^n states.

Theorem 2.2 *If L is DFA-regular then L^* is DFA-regular.*

Proof: L is DFA-regular via $M = (Q, \Sigma, \delta, s, F)$. We construct a DFA for L^*.

$$(2^Q, \Sigma, \delta', \{s\}, F')$$

where we define σ' and F' now.

$\delta'(T, \sigma) = \delta(T, \sigma)$ if $T \cap F = \emptyset$

$\delta'(T, \sigma) = \delta(T, \sigma) \cup \{\delta(s, \sigma)\}$ if $T \cap F \neq \emptyset$

$F' = \{T \mid T \cap F \neq \emptyset\}$.

∎

3 The Intersection of a CFG and a REG is CFG

It is well known that the intersection of a context free language and a regular language is context free. This theorem is used in several proofs that certain languages are not context free. The usual proof of this theorem is a cross product construction of a PDA and a DFA. This requires the equivalence of PDA's and CFG's. Is there a proof that does not use the equivalence? That is, is there a proof that just uses CFG's? There is and we show it in this note.

This proof is due to Y. Bar-Hillel et al. [1].

Def 3.1 A context free grammar is in *Chomsky Normal Form* if every production is either of the form $X \rightarrow YZ$ or $X \rightarrow \sigma$ where $\sigma \in \Sigma$.

The following lemmas are well known.

Lemma 3.2 *If L is a context free language without e then there is grammar in Chomsky Normal Form that generates L.*

Lemma 3.3 *If $L \neq \emptyset$ and L is regular then L is the union of regular language $A_1, \ldots, A_n$ where each A_i is accepted by a DFA with exactly one final state.*

We now prove our main theorem.

Theorem 3.4 *If L_1 is a context free language and L_2 is a regular language then $L_1 \cap L_2$ is context free.*

Proof:

We do the case where $e \notin L_1$ and $L_2 \neq \emptyset$. All other cases we leave to the reader.

By Lemma 3.2 we can assume there exists a Chomsky normal form grammar $G = (N, \Sigma, S, P)$ for L_1. By Lemma 3.3 $L_2 = A_1 \cup \cdots \cup A_n$ where each A_i where each A_i is recognized by a DFA with exactly one final state. Note that

$$L_1 \cap L_2 = L_1 \cap (A_1 \cup \cdots \cup A_n) = \bigcup_{i=1}^{n} (L_1 \cap A_i).$$

Since CFL's are closed under union (and this can be proven using CFG's, so this is not a cheat) we need only show that the intersection of L_1 with a regular language recognized by a DFA with one final state is CFL. Let $M = (Q, \Sigma, \delta, s, \{f\})$ be a DFA with exactly one final state.

We construct the CFG $G' = (N', \Sigma, S', P')$ for $L_1 \cap L(M)$.

1. The nonterminals N' are triples $[p, V, r]$ where $V \in N$ and $p, r \in Q$.

2. For each production $A \rightarrow BC$ in P, for every $p, q, r \in Q$ we have the production

   $$[p, A, r] \rightarrow [p, B, q][q, C, r]$$

   in P'.

3. For every production $A \rightarrow \sigma$ in P, for every $(p, \sigma, q) \in Q \times \Sigma \times Q$ such that $\delta(p, \sigma) = q$ we have the production

   $$[p, A, q] \rightarrow \sigma$$

   in P'

4. $S' = [s, S, f]$

We leave the easy proof that this works to the reader.

▮

References

[1] Y. Bar-Hiller, M. Perles, and E. Shamir. On formal properties of simple phrase structure grammars. *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationforshung*, 14(2):143–172, 1961.