CS 591 S2—Formal Language Theory: Integrating Experimentation and Proof—Fall 2018

Problem Set 3

Model Answers

Problem 1

First, we use Forlan to bind the identifier `ys` to the language $\{0,1\}^4 - \{\mathsf{0101}, \mathsf{1010}\}$, which has 14 elements:

```
- val xs = StrSet.power(StrSet.fromString "0, 1", 4);
val xs = - : str set
- val ys = StrSet.minus(xs, StrSet.fromString "0101, 1010");
val ys = - : str set
- Set.size ys;
val it = 14 : int
- StrSet.output("", ys);
0000, 0001, 0010, 0011, 0100, 0110, 0111, 1000, 1001, 1011, 1100, 1101, 1110,
1111
val it = () : unit
```

Next, we turn `ys` into a regular expression, `reg`. We then simplify `reg`, using first `SOME 100`, and then `SOME 1000`, as the limits on structural reorganizations (using `NONE` would run too long to obtain an answer) and `Reg.obviousSubset` as the conservative approximation to subset testing, binding the results to `reg'` and `reg''`, respectively. We then verify that `reg'` and `reg''` are equal, and check that `reg'` is locally simplified:

```
- val reg = Reg.fromStrSet ys;
val reg = - : reg
- val (full', reg') = Reg.locallySimplify (SOME 100, Reg.obviousSubset) reg;
val full' = false : bool
val reg' = - : reg
- val (full'', reg'') = Reg.locallySimplify (SOME 1000, Reg.obviousSubset) reg;
val full'' = false : bool
val reg'' = - : reg
- Reg.equal(reg', reg'');
val it = true : bool
- Reg.locallySimplified Reg.obviousSubset reg';
val it = true : bool
```

(Of course, this doesn't let us conclude that using a higher limit in the call to `locallySimplify` wouldn't have yielded a result with lower complexity.) Finally, we display `reg'` and print out its closure complexity, size, number of concatenations, number of symbols, and standardization status:

```
- Reg.output("", reg');
```

```
0(0(0 + 1)(0 + 1) + 1(00 + 1(0 + 1))) + 1(0(00 + (0 + 1)1) + 1(0 + 1)(0 + 1))
val it = () : unit
- Reg.ccToList(Reg.cc reg');
val it = [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] : int list
- length it;
val it = 24 : int
- Reg.size reg';
val it = 47 : int
- Reg.numConcats reg';
val it = 12 : int
- Reg.numSyms reg';
val it = 24 : int
- Reg.standardized reg';
val it = true : bool
```

Thus, our regular expression is

$$\begin{gathered} 0(0(0+1)(0+1)+1(00+1(0+1))) \\ +1(0(00+(0+1)1)+1(0+1)(0+1)). \end{gathered}$$

Our regular expression is correct by construction, but we can also check its correctness directly:

```
- StrSet.equal(Reg.toStrSet reg', ys);
val it = true : bool
```

Problem 2

(a) If **hasEmp**(α), i.e., $\% \in L(\alpha)$, then reduction rule (8) would also apply to $(\alpha\beta^*)^*$ (**hasEmp**(β^*) holds, because β^* is a closure), yielding $(\alpha + \beta^*)^*$. This can be restructured using structural rule (5) into $(\beta^* + \alpha)^*$, which can then be turned into $(\beta + \alpha)^*$ using reduction rule (7). So the rationale for restricting reduction rule (14) to only apply when **hasEmp**(α) is false, is to prefer the above sequence of simplifications, over one starting with reduction rule (14).

(If the conservative approximation of *sub* is powerful enough, then reduction rule (14) can be followed by reduction rules (1) and (4), also resulting in $(\alpha + \beta)^*$. But it's nicer not to have to depend upon properties of *sub*, or to incur the cost of running *sub*.)

(b) First we prove a lemma:

Lemma PS3.2.1
For all $ns, ms \in \mathbf{CC}$, *if* $ns <_{cc} ms$ *and* $|ns| \geq |ms|$, *then* $[0] \cup ns <_{cc} ms$.

Proof. Suppose $ns, ms \in \mathbf{CC}$, $ns <_{cc} ms$ and $|ns| \geq |ms|$. Because $ns <_{cc} ms$ and ns is at least as long as ms, there is an $i \in \mathbb{N} - \{0\}$ such that

- $i \leq |ns|$ and $i \leq |ms|$,
- for all $j \in [1 : i-1]$, $ns\,j = ms\,j$; and
- $ns\,i < ms\,i$.

Thus it follows that $i \in \mathbb{N} - \{0\}$,

- $i \leq |ns| < |ns| + 1 = |ns \mathbin{@} [0]|$ and $i \leq |ms|$;
- for all $j \in [1 : i - 1]$, $(ns \mathbin{@} [0])\, j = ns\, j = ms\, j$; and
- $(ns \mathbin{@} [0])\, i = ns\, i < ms\, i$.

Because 0 is $\leq$ every element of ns, we have that $[0] \cup ns = ns \mathbin{@} [0]$, and so we can conclude from the above that $[0] \cup ns <_{cc} ms$. $\square$

Suppose $\alpha, \beta \in \mathbf{Reg}$ and $\mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\beta} <_{cc} \overline{\overline{\mathbf{cc}\,\beta}}$. By Proposition 3.3.1, we have that

$$\begin{aligned}
\mathbf{cc}(\% + \alpha(\alpha + \beta)^*) &= \mathbf{cc}\,\% \cup \mathbf{cc}(\alpha(\alpha + \beta)^*) = [0] \cup \mathbf{cc}\,\alpha \cup \mathbf{cc}((\alpha + \beta)^*) \\
&= [0] \cup \mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}(\alpha + \beta)} = [0] \cup \mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\alpha \cup \mathbf{cc}\,\beta} \\
&= [0] \cup \mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\alpha} \cup \overline{\mathbf{cc}\,\beta}, \\
\mathbf{cc}((\alpha\beta^*)^*) &= \overline{\mathbf{cc}(\alpha\beta^*)} = \overline{\mathbf{cc}\,\alpha \cup \mathbf{cc}(\beta^*)} = \overline{\mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\beta}} = \overline{\mathbf{cc}\,\alpha} \cup \overline{\overline{\mathbf{cc}\,\beta}}.
\end{aligned}$$

Thus to show $\mathbf{cc}(\% + \alpha(\alpha + \beta)^*) <_{cc} \mathbf{cc}((\alpha\beta^*)^*)$, it will suffice to show $[0] \cup \mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\alpha} \cup \overline{\mathbf{cc}\,\beta} <_{cc} \overline{\mathbf{cc}\,\alpha} \cup \overline{\overline{\mathbf{cc}\,\beta}}$. By Propositions 3.3.1 and 3.3.3, it will suffice to show $[0] \cup \mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\beta} <_{cc} \overline{\overline{\mathbf{cc}\,\beta}}$. And this follows by Lemma PS3.2.1, since $\mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\beta} <_{cc} \overline{\overline{\mathbf{cc}\,\beta}}$ and $|\mathbf{cc}\,\alpha \cup \overline{\mathbf{cc}\,\beta}| \geq \overline{\mathbf{cc}\,\beta} = \overline{\overline{\mathbf{cc}\,\beta}}$.

(c) Suppose A and B are languages. To show that $(AB^*)^* = \{\%\} \cup A(A \cup B)^*$, it will suffice to show that $(AB^*)^* \subseteq \{\%\} \cup A(A \cup B)^* \subseteq (AB^*)^*$.

$((AB^*)^* \subseteq \{\%\} \cup A(A \cup B)^*)$ It will suffice to show that, for all $n \in \mathbb{N}$, $(AB^*)^n \subseteq \{\%\} \cup A(A \cup B)^*$.

- **(basis step)** We have that $(AB^*)^0 = \{\%\} \subseteq \{\%\} \cup A(A \cup B)^*$.
- **(inductive step)** Suppose $n \in \mathbb{N}$, and assume the inductive hypothesis: $(AB^*)^n \subseteq \{\%\} \cup A(A \cup B)^*$. We must show that $(AB^*)^{n+1} \subseteq \{\%\} \cup A(A \cup B)^*$. By the inductive hypothesis, we have that $(AB^*)^{n+1} = (AB^*)(AB^*)^n \subseteq (AB^*)(\{\%\} \cup A(A \cup B)^*)$. Thus it will suffice to show that $(AB^*)(\{\%\} \cup A(A \cup B)^*) \subseteq \{\%\} \cup A(A \cup B)^*$. Suppose $w \in (AB^*)(\{\%\} \cup A(A \cup B)^*)$, so that $w = xy$ for some $x \in AB^*$ and $y \in \{\%\} \cup A(A \cup B)^*$. We must show that $w \in \{\%\} \cup A(A \cup B)^*$. There are two subcases to consider.
  - Suppose $y = \%$. Thus $w = xy = x\% = x \in AB^*$. But $B \subseteq A \cup B$, and thus $B^* \subseteq (A \cup B)^*$. Thus $w \in AB^* \subseteq A(A \cup B)^* \subseteq \{\%\} \cup A(A \cup B)^*$.
  - Suppose $y \in A(A \cup B)^*$. Since $A \subseteq A \cup B$, it follows that $A \subseteq A^* \subseteq (A \cup B)^*$. Since $B \subseteq A \cup B$, it follows that $B^* \subseteq (A \cup B)^*$. Thus $w = xy \in AB^*A(A \cup B)^* \subseteq A(A \cup B)^*(A \cup B)^*(A \cup B)^* = A(A \cup B)^*(A \cup B)^* = A(A \cup B)^* \subseteq \{\%\} \cup A(A \cup B)^*$.

$(\{\%\} \cup A(A \cup B)^* \subseteq (AB^*)^*)$ Because $\% \in (AB^*)^*$, it will suffice to show that $A(A \cup B)^* \subseteq (AB^*)^*$. Let $X = \bigcup\{\, (AB^*)^{n+1} \mid n \in \mathbb{N} \,\}$. Since $X \subseteq (AB^*)^*$, it will suffice to show that, for all $n \in \mathbb{N}$, $A(A \cup B)^n \subseteq X$. (If $w \in A(A \cup B)^*$, then $w = xy$ for some $x \in A$ and $y \in (A \cup B)^*$, so that $y \in (A \cup B)^n$ for some $n \in \mathbb{N}$. Thus $w = xy \in A(A \cup B)^n \subseteq X \subseteq (AB^*)^*$.) We proceed by mathematical induction.

- **(basis step)** We have that $A(A \cup B)^0 = A\{\%\} \subseteq AB^* \subseteq (AB^*)^1 = (AB^*)^{0+1} \subseteq X$.
- **(inductive step)** Suppose $n \in \mathbb{N}$, and assume the inductive hypothesis: $A(A \cup B)^n \subseteq X$. We must show that $A(A \cup B)^{n+1} \subseteq X$. By the inductive hypothesis, we have that $A(A \cup B)^{n+1} = A(A \cup B)^n(A \cup B) \subseteq X(A \cup B) = XA \cup XB$. So it remains to show that $XA \cup XB \subseteq X$, and we do this by showing that $XA \subseteq X$ and $XB \subseteq X$.

  To see that $XA \subseteq X$, suppose $w \in XA$, so that $w = xy$ for some $x \in X$ and $y \in A$. Thus $x \in (AB^*)^{n+1}$ for some $n \in \mathbb{N}$. We have that $A = A\{\%\} \subseteq AB^* = (AB^*)^1$. Thus $w = xy \in (AB^*)^{n+1}A \subseteq (AB^*)^{n+1}(AB^*)^1 = (AB^*)^{(n+1)+1} \subseteq X$.

  To see that $XB \subseteq X$, suppose $w \in XB$, so that $w = xy$ for some $x \in X$ and $y \in B$. Thus $x \in (AB^*)^{n+1}$ for some $n \in \mathbb{N}$. Since $(AB^*)^{n+1} = (AB^*)^n(AB^*) = (AB^*)^n AB^*$, we have that $x = tuv$ for some $t \in (AB^*)^n$ and $u \in A$ and $v \in B^*$. Thus $w = xy = tuvy \in (AB^*)^n AB^*B \subseteq (AB^*)^n AB^* = (AB^*)^n(AB^*)^1 = (AB^*)^{n+1} \subseteq X$.

(d) Suppose $\alpha, \beta \in \mathbf{Reg}$. By part (c), we have that

$$\begin{aligned} L((\alpha\beta^*)^*) &= L(\alpha\beta^*)^* = (L(\alpha)\, L(\beta^*))^* = (L(\alpha)\, L(\beta)^*)^* = \{\%\} \cup L(\alpha)(L(\alpha) \cup L(\beta))^* \\ &= \{\%\} \cup L(\alpha)\, L(\alpha+\beta)^* = \{\%\} \cup L(\alpha)\, L((\alpha+\beta)^*) = L(\%) \cup L(\alpha(\alpha+\beta)^*) \\ &= L(\% + \alpha(\alpha+\beta)^*). \end{aligned}$$

Thus $(\alpha\beta^*)^* \approx \% + \alpha(\alpha+\beta)^*$.

Problem 3

(a)

$$(1 + 0^*11)^*0^*(\% + 1)$$

(b) Let

$$Y = \{1\} \cup \{0\}^*\{11\} \qquad \text{and} \qquad Z = \{0\}^*\{\%, 1\}.$$

Since $L(\alpha) = Y^*Z$, it will suffice to show that $Y^*Z = X$. We begin by proving some lemmas.

Let $U = \{\, w \in X \mid \text{neither } 0 \text{ nor } 01 \text{ is a suffix of } w \,\}$.

Lemma PS3.3.1
$UU \subseteq U$.

Proof. Suppose $w \in UU$, so that $w = xy$, for some $x, y \in U$. We must show that $w \in U$. Since $U \subseteq X \subseteq \{0,1\}^*$, we have that $w \in \{0,1\}^*$.

Suppose, toward a contradiction, that $w \notin X$. Thus 010 is a substring of $w = xy$. Because $x, y \in U$, we have that $x, y \in X$. Thus there are two cases to consider.

- (0 is a suffix of x and 10 is a prefix of y) But this contradicts the fact that $x \in U$.
- (01 is a suffix of x and 0 is a prefix of y) But this contradicts the fact that $x \in U$.

Since we obtained a contradiction in both cases, we have an overall contradiction. Thus $w \in X$.

Suppose, toward a contradiction, that $\mathsf{0}$ is a suffix of w. Since $w = xy$, there are two cases to consider.

- ($\mathsf{0}$ is a suffix of y) But this contradicts the fact that $y \in U$.
- ($\mathsf{0}$ is a suffix of x and $y = \%$) But this contradicts the fact that $x \in U$.

Since we obtained a contradiction in both cases, we have an overall contradiction. Thus $\mathsf{0}$ is not a suffix of w.

Suppose, toward a contradiction, that $\mathsf{01}$ is a suffix of w. Since $w = xy$, there are three cases to consider.

- ($\mathsf{01}$ is a suffix of y) But this contradicts the fact that $y \in U$.
- ($\mathsf{0}$ is a suffix of x, and $y = \mathsf{1}$) But this contradicts the fact that $x \in U$.
- ($\mathsf{01}$ is a suffix of x, and $y = \%$) But this contradicts the fact that $x \in U$.

Since we obtained a contradiction in all three cases, we have an overall contradiction. Thus $\mathsf{01}$ is not a suffix of w.

Since $w \in X$, and neither $\mathsf{0}$ nor $\mathsf{01}$ is a suffix of w, we have that $w \in U$. □

Lemma PS3.3.2
$Y \subseteq U$.

Proof. Suppose $w \in Y$. We must show that $w \in U$. There are two cases to consider.

- ($w = \mathsf{1}$) Since $\mathsf{1} \in X$ and neither $\mathsf{0}$ nor $\mathsf{01}$ is a suffix of $\mathsf{1}$, we have that $w = \mathsf{1} \in U$.
- ($w \in \{\mathsf{0}\}^*\{\mathsf{11}\}$) Hence $w = \mathsf{0}^n\mathsf{11}$, for some $n \in \mathbb{N}$. Since even $\mathsf{10}$ isn't a substring of w, we have that $w \in X$. And, since neither $\mathsf{0}$ nor $\mathsf{01}$ is a suffix of $\mathsf{11}$, we have that neither $\mathsf{0}$ nor $\mathsf{01}$ is a suffix of w. Thus $w \in U$.

□

Lemma PS3.3.3
$Y^* \subseteq U$.

Proof. It will suffice to show that, for all $n \in \mathbb{N}$, $Y^n \subseteq U$. We proceed by mathematical induction.

(Basis) Since $\% \in U$, we have that $Y^0 = \{\%\} \subseteq U$.

(Inductive Step) Suppose $n \in \mathbb{N}$, and assume the inductive hypothesis: $Y^n \subseteq U$. We must prove that $Y^{n+1} \subseteq U$. We have that

$$\begin{aligned} Y^{n+1} &= YY^n \\ &\subseteq UU && \text{(Lemma PS3.3.2 and the inductive hypothesis)} \\ &\subseteq U && \text{(Lemma PS3.3.1).} \end{aligned}$$

□

Lemma PS3.3.4
$Y^*Z \subseteq X$.

Proof. By Lemma PS3.3.3, we have that $Y^*Z \subseteq UZ$. Thus, it will suffice to show that $UZ \subseteq X$. Suppose $w \in UZ$, so that $w = xy$, for some $x \in U$ and $y \in Z$. Because $U \subseteq \{0,1\}^*$ and $Z \subseteq \{0,1\}^*$, we have that $w \in \{0,1\}^*$. We must show that $w \in X$. Suppose, toward a contradiction, that $w \notin X$. Thus 010 is a substring of $w = xy$. Since $y \in Z$, there are two cases to consider.

- ($y = 0^n$, for some $n \in \mathbb{N}$) Because 010 is a substring of $w = xy$, and $x \in X$, we have that 01 is a suffix of x. But this contradicts the fact that $x \in U$.
- ($y = 0^n1$, for some $n \in \mathbb{N}$) Because 010 is a substring of $w = xy$, and $x \in X$, we have that 01 is a suffix of x. But this contradicts the fact that $x \in U$.

Since we obtained a contradiction in each case, we have an overall contradiction. Thus $w \in X$. □

Lemma PS3.3.5
$X \subseteq Y^*Z$.

Proof. Since $X \subseteq \{0,1\}^*$, it will suffice to show that, for all $w \in \{0,1\}^*$,

$$\text{if } w \in X, \text{ then } w \in Y^*Z.$$

We prove this using strong string induction. Suppose $w \in \{0,1\}^*$, and assume the inductive hypothesis: for all $x \in \{0,1\}^*$, if $|x| < |w|$, then

$$\text{if } x \in X, \text{ then } x \in Y^*Z.$$

We must show that

$$\text{if } w \in X, \text{ then } w \in Y^*Z.$$

Suppose $w \in X$. We must show that $w \in Y^*Z$. There are three cases to consider.

- ($w = \%$) Then $w = \%(\%\%) \in Y^*Z$.
- ($w = 1x$, for some $x \in \{0,1\}^*$) Since x is a substring of w, we have that $x \in X$. Thus, by the inductive hypothesis, we have that $x \in Y^*Z$. Hence $w = 1x \in Y(Y^*Z) = (YY^*)Z \subseteq Y^*Z$.
- ($w = 0x$, for some $x \in \{0,1\}^*$) Let $n \in \mathbb{N}$ be largest such that 0^n is a prefix of x (n is well-defined, because it could always be 0, and can't be bigger than $|x|$); let $y \in \{0,1\}^*$ be such that $x = 0^n y$. Thus $w = 0^{n+1}y$, and 0 is not a prefix of y.

  If $y = \%$, then $w = 0^{n+1} = \%(0^{n+1}\%) \in Y^*Z$. So, suppose $y \neq \%$, so that $y = 1z$ for some $z \in \{0,1\}^*$. Thus $w = 0^{n+1}1z$.

  If $z = \%$, then $w = 0^{n+1}1 = \%(0^{n+1}1) \in Y^*Z$. So, suppose $z \neq \%$. But 0 is not a prefix of z, since then $0^n(010) = 0^{n+1}10$ would be a prefix of w. So, $z = 1u$, for some $u \in \{0,1\}^*$. Thus $w = 0^{n+1}11u$. Since u is a substring of w, we have that $u \in X$. Thus, by the inductive hypothesis, it follows that $u \in Y^*Z$. Hence $w = (0^{n+1}(11))u \in Y(Y^*Z) = (YY^*)Z \subseteq Y^*Z$.

□

By Lemmas PS3.3.4 and PS3.3.5, we have that $Y^*Z \subseteq X \subseteq Y^*Z$, i.e., $Y^*Z = X$.