Problem Set 3: Week 3: Theory of Lexical and Syntactic Analysis Solution Key

C S 395T

15 September 2021

This is the problem set for week 3 of C S 395T SIMPL. It is intended to help you learn the material by working out more examples and exercises than is possible to cover in the videos. Feel free to work individually or in groups. Ask questions on Piazza. You are not required to submit anything, and the problem set doesn't count directly towards your course grade. The solution key for the problem set will be made available one week after its release.

Problem 1. *Finite automata and regular expressions.*

Construct both non-deterministic and deterministic finite automata recognizing the language denoted by the regular expression $(a|b)^*abb(a|b)^*$. Show the sequence of moves made by each automaton in processing the input string $ababbab$.

Solution: The solutions are not unique, but a 4-state DFA or a 4-state NFA suffices to recognize this language. The 4-state DFA looks like this: $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{a, b\}$, $S = q_0$, $F = q_3$, $\delta = \{(q_0, a) \to q_1, (q_0, b) \to q_0, (q_1, a) \to q_1, (q_1, b) \to q_2, (q_2, a) \to q_1, (q_2, b) \to q_3, (q_3, a) \to q_3, (q_3, b) \to q_3\}$.

The sequence of moves made by each automaton is left as an exercise for the student.

Problem 2. *Regular expressions in Ruby.*

Use the websites `https://rubular.com` and `https://ruby-doc.org/core-3.0.2/Regexp.html` to understand the syntax of regular expressions in Ruby.

(a) Give an example of a string that matches the regular expression `/^\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}$/` and another that doesn't match it.

(b) You have the following string:

```
"Hughes Missile Systems Company, Tucson, Arizona, is being awarded
a $7,311,983 modification to a firm fixed price contract for the FY94
TOW missile production buy, total 368 TOW 2Bs. Work will be performed
in Tucson, Arizona, and is expected to be completed by April 30, 1996.
Of the total contract funds, $7,311,983 will expire at the end of the
current fiscal year. This is a sole source contract initiated on January 14,
1991. The contracting activity is the U.S. Army Missile Command,
Redstone Arsenal, Alabama (DAAH01-92-C-0260).

Conventional Munitions Systems, Incorporated, Tampa, Florida, is being
awarded a $6,952,821 modification to a firm fixed price contract for
Dragon Safety Circuits Installation and retrofit of Dragon I Missiles with
Dragon II Warheads. Work will be performed in Woodberry, Arkansas
(90%), and Titusville, Florida (10%), and is expected to be completed by
```

```
May 31, 1996. Contract funds will not expire at the end of the current
fiscal year. This is a sole source contract initiated on May 2, 1994. The
contracting activity is the U.S. Army Missile Command, Redstone
Arsenal, Alabama (DAAH01-94-C-S076)."
```

What are the matches that the regular expressions
`/\$[\d,]+/`,
`/\w+ \d{1,2}, \d{4}/`,
and `(?:\s*\b([A-Z]+(?:\s*\w*)?)\b)+`
will show in this string?

Solution:

(a) Matching: `128.2.254.999`. Not matching: `128.2.254` or `1729.48.762.100`.

(b) The matches for regexp `/\$[\d,]+/` are `$7,311,983` and `6,952,821`.
The matches for regexp `/\w+ \d{1,2}, \d{4}/` are `April 30, 1996`, `January 14, 1991`, `May 31, 1996`, and `May 2, 1994`.
The matches for regexp `(?:\s*\b([A-Z]+(?:\s*\w*)?)\b)+` are `Hughes Missile Systems Company`, `Tucson`, `Arizona`, and several more.
See `http://ruby.bastardsbook.com/chapters/regexes/` for further details.

Problem 3. *Context-free grammars.*
Consider the following grammar:
$\langle S \rangle \rightarrow$ `(` $\langle L \rangle$ `)`
$\langle S \rangle \rightarrow$ `a`
$\langle L \rangle \rightarrow \langle L \rangle$ `,` $\langle S \rangle$
$\langle L \rangle \rightarrow \langle S \rangle$

(a) What are the terminals, non-terminals, and start symbol?

(b) Construct the parsing table for the grammar.

(c) Find a parse tree for the sentence `(a, (a, a))`.

(d) Construct a leftmost and a rightmost derivation for the sentence `(a, (a, a))`.

(e) Construct the GFG for this grammar. Show the steps of the NGA in parsing the sentence `(a, (a, a))`.

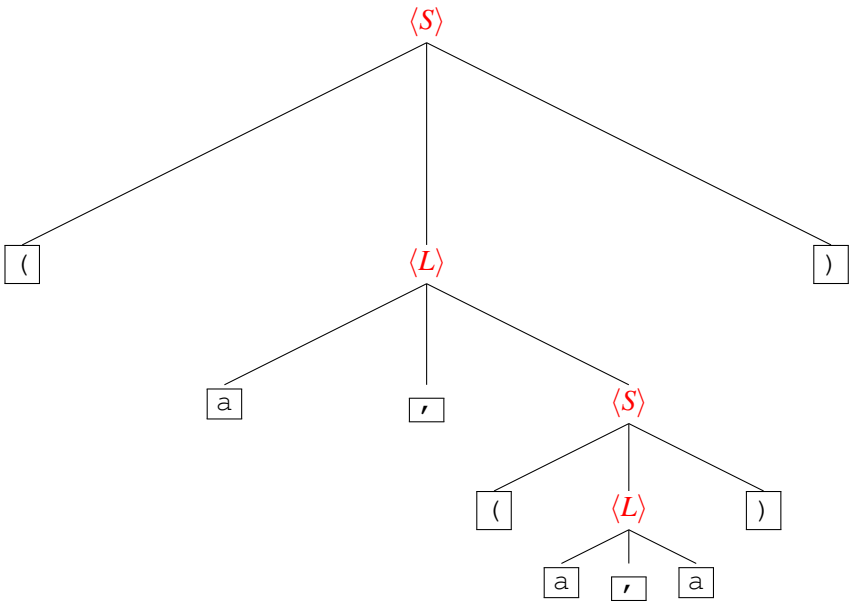
Solution:

(a) The terminals are `(`, `)`, `a`, `,`, and the implicit `$` for end-of-input. The non-terminals are $\langle S \rangle$ and $\langle L \rangle$. The start symbol is $\langle S \rangle$.

I realize that the slides say that the start symbol can't appear on the right side of a production. This is a normalization condition that is imposed on CFGs for cleaning up theoretical presentations. It is not critical and can be easily fixed if needed: simply add an extra production $\langle Start \rangle \rightarrow \langle S \rangle$ and make $\langle Start \rangle$ be the start symbol.

(b) Left as an exercise for the student.

(c) The parse tree is



I have simplified the tree somewhat by removing chains of replacements.

(d) Leftmost derivation:

$$
\begin{aligned}
\langle S\rangle &\Longrightarrow (\,\langle L\rangle\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,\langle S\rangle\,) \\
&\Longrightarrow (\,\langle S\rangle\,,\,\langle S\rangle\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,\langle S\rangle\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\langle L\rangle\,)\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\langle L\rangle\,,\,\langle S\rangle\,)\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\langle S\rangle\,,\,\langle S\rangle\,)\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\mathtt{a}\,,\,\langle S\rangle\,)\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\mathtt{a}\,,\,\mathtt{a}\,)\,).
\end{aligned}
$$

Rightmost derivation:

$$
\begin{aligned}
\langle S\rangle &\Longrightarrow (\,\langle L\rangle\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,\langle S\rangle\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,(\,\langle L\rangle\,)\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,(\,\langle L\rangle\,,\,\langle S\rangle\,)\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,(\,\langle L\rangle\,,\,\mathtt{a}\,)\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,(\,\langle S\rangle\,,\,\mathtt{a}\,)\,) \\
&\Longrightarrow (\,\langle L\rangle\,,\,(\,\mathtt{a}\,,\,\mathtt{a}\,)\,) \\
&\Longrightarrow (\,\langle S\rangle\,,\,(\,\mathtt{a}\,,\,\mathtt{a}\,)\,) \\
&\Longrightarrow (\,\mathtt{a}\,,\,(\,\mathtt{a}\,,\,\mathtt{a}\,)\,).
\end{aligned}
$$

(e) Left as an exercise for the student.