CS323 Compilers Assignment1

12110644 周思呈

Exercise 1

Java programs may contain lexical errors. Please give at least two types of possible lexical errors in Java programs and provide code snippets as examples when possible. [10 points]

1. Identifier Errors. Identifier should not start with a digit and should avoid special characteristics.

```
int 3a = 6; // identifier starts with a digit
```

2. Unclosed Comment. Comments should begin with `/*` and end with `*/`.

```
/* This is an unclosed comment
```

Exercise 2

Given a string s, can you find a string x that is both a prefix and a suffix of s? Can you further find a string y that is both a proper prefix and a proper suffix of s? If yes, please provide an example. Otherwise, please explain the reason. [10 points]

1. String s itself is both a prefix and suffix of s.

2. String y can't be both a proper prefix and a proper suffix of s. A proper prefix should include the first letter of x and uninclude the last letter, while a proper suffix of s should include the last letter and uninclude the first letter. It can not be done at the same time.

Exercise 3

In a string of length n (n > 0), how many of the following are there? For simplicity, we assume that the string contains n different characters. Besides giving the final answers, please also explain how you derive the answers.

1. **Substrings of length m (0 < m ≤ n) [15 points]**

$$m - n + 1 \tag{1}$$

2. **Subsequences [15 points]**

$$2^n \tag{2}$$

Exercise 4

Write a regular definition as well as a regular expression to represent all strings of valid telephone numbers in Shenzhen. A valid telephone number contains the country code (86), a hyphen, the area code 755, another hyphen, and eight digits where the first one cannot be zero (e.g., 86-755-88015159). [10 points]

Exercise 5

Given an alphabet Σ = {0, 1}, are the following two regular languages equivalent? Besides saying yes or no, please also prove your answer. [20 points]

1. $L_1 = L((0^*1^*)^*)$
2. $L_2 = L((0|1)^*)$

They are equivalent.

$L_1 = L((0^*1^*)^*) = L((0^*1^*)^{**}) = L_1^*$ (because * is idempotent)

Word in $L(0|1)$ is either 0 or 1, which can be represented by L_1. So $L_2 \subseteq L_1$.

Similarly, $L_2 = L((0|1)^{**}) = L_2^*$. Every word in $L(0^*1^*)$ can be represented by L_2. So $L_1 \subseteq L_2$.

Exercise 6

Consider the regular expression ba∗|ab∗. Please provide a state transition diagram that can recognize the strings in the corresponding regular language. Can the transition diagram recognize the string baab? If yes, please give the sequence of state transitions. Otherwise, please explain the reason. [20 points]
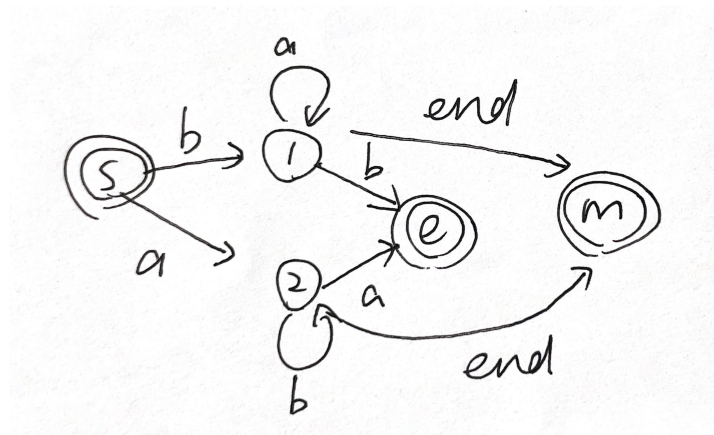
String "baab" can not be recognized. The sequence is start -> node 1 -> node 1 -> node 1 -> end(fail).