



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COMPILIER @BY 2019

for SUSTech CSE

HOMEWORK 1

EDITED BY

汪至圆

11610634

2019
SHENZHEN

1 Exercise 1: In a string of length n ($n > 0$), how many of the following are there?

1.1 Prefixes [5 points]

There are $n+1$ prefixes.

1.2 Suffixes [5 points]

There are $n+1$ suffixes.

1.3 Proper prefixes [5 points]

There are $n-1$ proper prefixes.

1.4 Prefixes of length m ($0 < m \leq n$) [5 points]

The number of prefixes of length m is 1.

1.5 Suffixes of length m ($0 < m \leq n$) [5 points]

The number of suffixes of length m is 1.

1.6 Proper prefixes of length m ($0 < m \leq n$) [5 points]

The number of proper prefixes of length m is 1 if $m < n$, is 0 if $m = n$.

1.7 Substrings [10 points]

There are $\frac{(n+1)n}{2} + 1$ substrings

1.8 Subsequences [10 points]

∴ For a subsequence of a string. Each char of the string just have two status, in the sequence or not in it.

∴ The number of the status for all the subsequences is: 2^n

∴ There are 2^n subsequences

2 Exercise 2: Describe the languages denoted by the following regular expressions:

2.1 $a(a|b)^*b$ [5 points]

The string start with a , there are n a or b behind it (n can be 0), then the string end with b .

$$s = a + n \times (a \text{ or } b) + b \quad (n \geq 0)$$

2.2 $((\epsilon|a)^*b^*)^*[5points]$

A pattern is start with n a, there are m of b behind it. The string consist of k pattern(m,n,k can be zero).

$$s = k \times (n \times (\text{null or } a) + m \times b) \quad (m \geq 0, n \geq 0, k \geq 0)$$

2.3 $(a|b)^*a(a|b)(a|b)[5points]$

A pattern consist of a or b. The string is start with n pattern, behind it is a, then end with 2 pattern.

$$\begin{aligned} p &= a \text{ or } b \\ s &= n \times p + a + 2 \times p \quad (n \geq 0) \end{aligned}$$

2.4 $a^*ba^*ba^*ba^*[5points]$

The pattern consist of n a and b, the string is consist of 3 pattern and b

$$\begin{aligned} p &= n \times a + b \\ s &= 3 \times p + m \times a \quad (n \geq 0, m \geq 0) \end{aligned}$$

3 Exercise 3: Write regular definitions for the following languages. Please provide brief explanations why your definitions are correct.

3.1 All strings representing valid telephone numbers in Shenzhen. A valid telephone number contains the area code 755 followed by eight digits where the first one cannot be zero (e.g., 75588015159). [10 points]

1 | `755[1-9][0-9]{7}`

3.2 All strings of uppercase letters (A-Z) in which the letters are in ascending lexicographic order. For example, “ABE” is such a string but “BAE” is not (because the letter B appears after A in the alphabet). [10 points]

1 | `A*B*C*D*E*F*G*H*I*J*K*L*M*N*O*P*Q*R*S*T*U*V*W*X*Y*Z*`

3.3 All strings of lowercase letters that contain the five vowels in order. [10 points]

1 | `p = [b-df-hj-np-tv-z]`
2 | `re = (a|{p})*a(e|{p})*e(i|{p})*i(o|{p})*o(u|{p})*({p})*`

4 Optional Exercises (20 bonus points)

4.1 Exercise 1: Write regular definitions for the following languages. Please provide brief explanations why your definitions are correct

4.1.1 Comments consisting of a string surrounded by `/*` and `*/`, without an intervening `*/`, unless it is inside double-quotes `""`. [10 points]

```
1 | /\*((\"\\*/\")|(?!\\"/>
```

4.1.2 All strings of digits with no repeated digits. [10 points]

```
1 | //If we want to get the longest left substring with no digits from the string:
2 | ((?<n>[0-9])(?!\\k<n>))*([0-9])
3 | //If we want to check the string have no repeat digits(If have repeat digits, we
   | won't get anything)
4 | ^((?<n>[0-9])(?!\\k<n>))*([0-9])$
```