

编译原理 - 作业(1) : 词法分析

Due: Mar 16, 2021 at the beginning of the class (14:20)

Exercise 3.1.2 (P114): Tagged languages like HTML or XML are different from conventional programming languages in that the punctuation (tags) are either very numerous (as in HTML) or a user-defined set (as in XML). Further, tags can often have parameters. Suggest how to divide the following HTML document:

Here is a photo of my house:

<P>

See More Pictures if you liked that one.<P>

into appropriate lexemes. Which lexemes should get associated lexical values, and what should those values be?

分为以下几类:

文本: "Here is a photo of", "my house", "See", "More Pictures", "if you liked that one".

标签: B, P, BR, IMG, SRC, A, HREF

标点符号: /, <, >, =

Exercise 3.3.5 (P125): Write regular definitions for the following languages:

- b). All strings of lowercase letters in which the letters are in ascending lexicographic order (词典递增序排列).
- c). Comments, consisting of a string surrounded by /* and */, without an intervening */, unless it is inside double-quotes ("").
- h). All strings of a's and b's that do not contain the substring abb.

b) $a^*b^*\dots z^*$

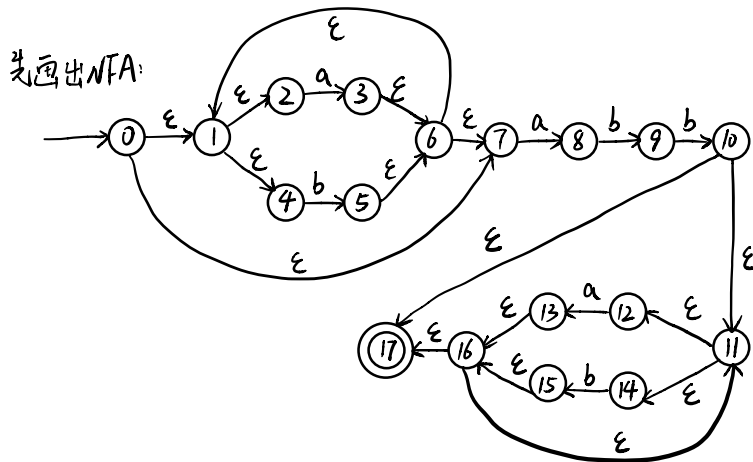
c) $\backslash / \backslash * ((\backslash * \backslash /) | (" * / ") | (. *)) * \backslash * \backslash /$

代表 /, 前面的 \ 为转义字符. 同理, 代表 *. 不含 */ 字符. 含 */ 字符时要在引号里面. 代表字符串, 如注释符等等. 取闭包运算. 同理, 代表 */.

h) $b^*(a|ba)^*b?$ 其中 $b?$ 代表 0 或 1 个 b. 这样在匹配到 a 时引检测后面只有 0 或 1 个 b, 不会出现 "abb" 型子串.

Exercise 3.7.3 (P166): Convert the following regular expressions to DFAs:

d). $(a|b)^*abb(a|b)^*$



然后列表:

	a	b
S_0	S_1	S_2
S_1	S_1	S_3
S_2	S_1	S_2
S_3	S_1	S_4
S_4	S_5	S_6
S_5	S_5	S_6
S_6	S_5	S_6

其中:

$$S_0 = \{0, 1, 2, 4, 7\}$$

$$S_1 = \{1, 2, 3, 4, 6, 7, 8\}$$

$$S_2 = \{1, 2, 4, 5, 6, 7\}$$

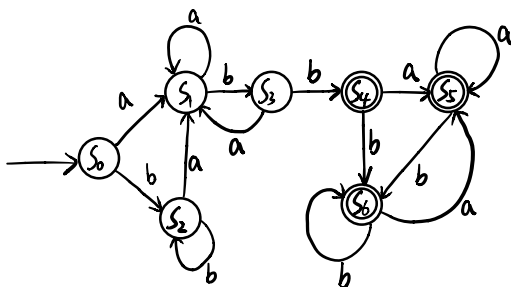
$$S_3 = \{1, 2, 4, 5, 6, 7, 9\}$$

$$S_4 = \{1, 2, 4, 5, 6, 7, 10, 11, 12, 14, 17\}$$

$$S_5 = \{1, 2, 3, 4, 6, 7, 8, 11, 12, 13, 14, 16, 17\}$$

$$S_6 = \{1, 2, 4, 5, 6, 7, 11, 12, 14, 15, 16, 17\}$$

按表绘制DFA如下:



Minimize the DFA as follows:

