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 <B>my house</B>:
<P><IMG SRC = “house.gif”>
See <A HREF = “morePix.html”>More Pictures</A> if you liked that one.<P>

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

<文本, Here is a photo of >, <开始标签, <B>>,
<文本, my house>, <结束标签, </B>>, <开始标签, P>.
<空元素, <IMG SRC = "house.gif">>, <空元素,
>,
<文本, See>, <开始标签, <A HREF = "morePix.html">>,
<文本, More Pictures>, <结束标签, </A>>,
<文本, if you liked that one.>, <结束标签, <P>>

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* c* d* e* f* g* h* i* j* k* l* m* n* * p* q* r* s* t* u* v* w* x* y* z*

c) \/*([^*"]|*[^\/"]|"[^"]*")**\/

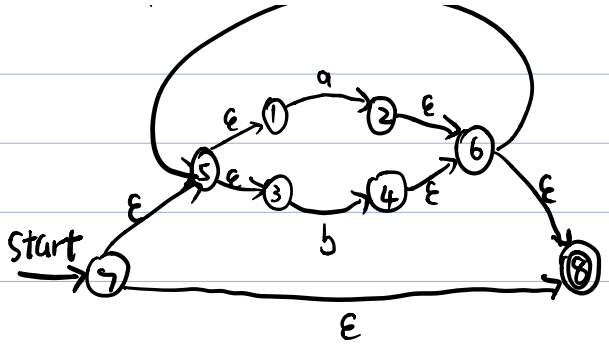
能匹配0以上次数* 为了不匹配/*""*/"*/ 为了匹配/**/

h) b*(a+b?)* => 因为连续b前不能是a，所以只能在开头

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

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

(a|b)*: NFA: ε



∴ DFA:

	a	b
13578	123568	124568
123568	123568	124568
124568	123568	124568

∵ 123568, 124568 同组且无法区分

设 13578 = A', 123568 = 124568 = B'

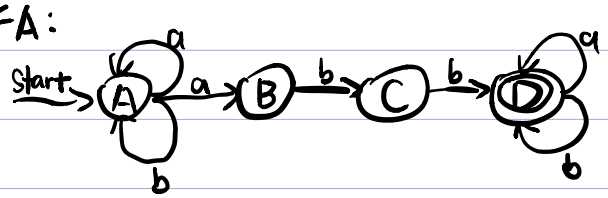
	a	b
A'	B'	B'
B'	B'	B'

A', B' 同组且无法区分

设 A'' = A' = B'

	a	b
A''	A''	A''

∴ 原题 NFA:



∴ DFA:

	a	b
A	AB	A
AB	AB	AC
AC	AB	AD

AD	ABD	AD
ABD	ABD	ACD
ACD	ABD	AD

∵ AD, ACD 同组，且不可区分

∴ 令 E = AD = ACD

∴

	a	b
A	AB	A
AB	AB	AC
AC	AB	E
E	ABD	E
ABD	ABD	E

∵ ABD, E 同组且不可区分

∴ 令 F = E = ABD

	a	b
A	AB	A
AB	AB	AC
AC	AB	F
F	F	F

∴ NFA: