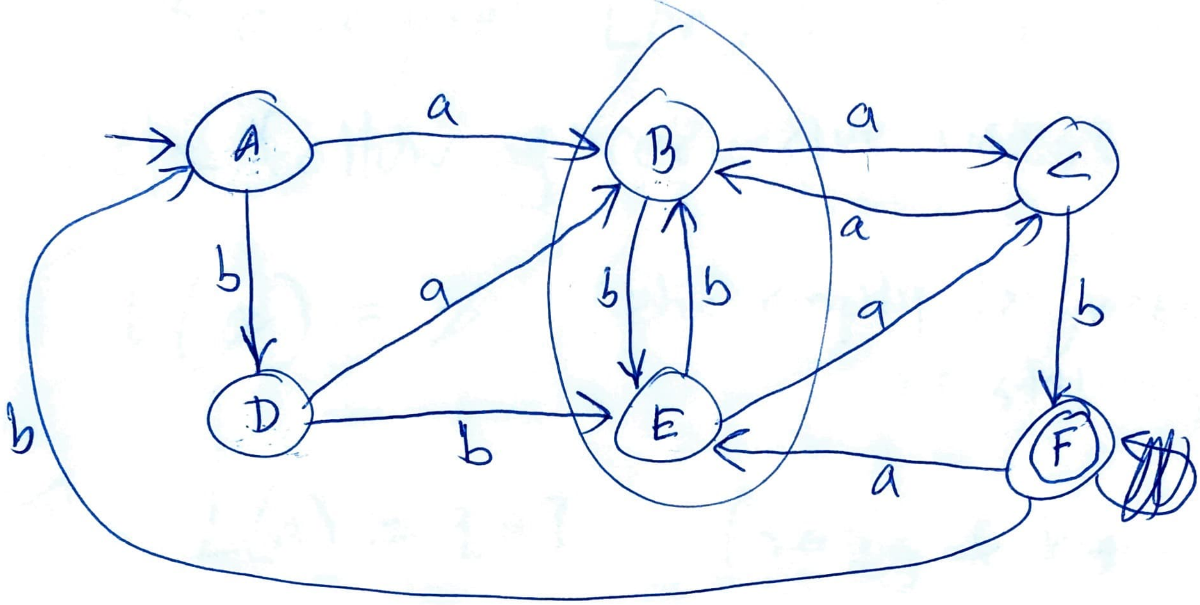
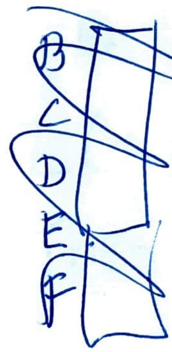
① Another DFA min example:

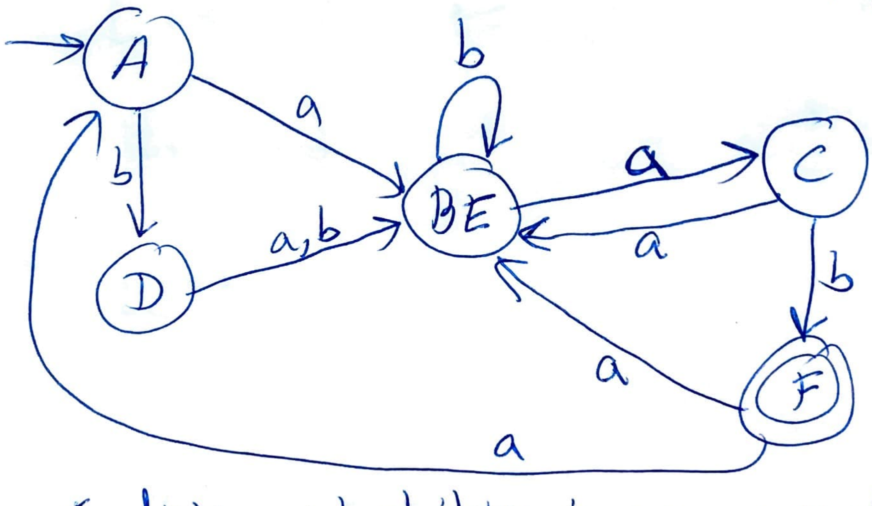
CSCE 355
2/2/22



Merge B & E
(only pair)



	A	B	C	D	E
B	X				
C	X	X			
D	X	X	X		
E	X	O	X	X	
F	X	X	X	X	X



Indistinguishability is an equivalance relation
(reflexive, symmetric, transitive)
equivalence classes are the states of the min DFA.

(2) <u>Def</u>: A regex r over an alphabet Σ <u>denotes</u> a language $L(r) \subseteq \Sigma^*$ according to the following recursive rules:

Base case

$L(\emptyset) := \emptyset$ (the empty language, no strings)

$a \in \Sigma$, $L(a) := \{a\}$ (string of length 1)

Let s and t be regexes over Σ and assume that $L(s)$ and $L(t)$ have already been defined. Then define

$$L(s+t) := L(S) \cup L(t)$$

~~L(st)~~

$$= \{w \in \Sigma^* : w \in L(s) \text{ or } w \in L(t)\}$$

(or both)

~~L(s)L(t)~~)

$$L(st) := L(s)L(t) := \{xy : x \in L(s) \;\&\; y \in L(t)\}$$

$$L(s^*) := \{\varepsilon\} \cup L(s) \cup L(s)L(s) \cup L(s)L(s)L(s) \cup \dots$$

$$:= \{w_1 \cdots w_n : n \geq 0 \;\&\; \forall i, 1 \leq i \leq n, w_i \in L(s)\}$$

---

3. $\Sigma = \{a, b, c\}$

$L(\underline{ab + c^*}) = \{$

$\cancel{L(a)}\ L(ab) = L(a)L(b) = \{a\}\{b\} = \{ab\}$

$L(c^*) = \{\varepsilon, c, cc, ccc, \ldots, \underbrace{c^n}_{\underbrace{c \cdots c}_{n \text{ times}}}, \ldots\}$

$L(c) = \{c\}$

$L(ab + c^*) = L(ab) \cup L(c^*)$

$= \{ab\} \cup \{\varepsilon, c, cc, \ldots\}$

$= \{ab, \varepsilon, c, cc, \ldots\}$

$L((ab)^*) = \{\varepsilon, ab, abab, ababab, \ldots\}$

$L(ab^*) = \underbrace{\{a\}}_{L(a)} \underbrace{\{\varepsilon, b, bb, bbb, \ldots\}}_{L(b^*)}$

$= \{a, ab, abb, abbb, \ldots\}$

$L((a+b)(a+b)) = L(a+b)L(a+b)$

$= (L(a) \cup L(b))(L(a) \cup L(b))$

$= \{a, b\}\{a, b\} = \{aa, ab, ba, bb\}$

4

$L((a+b)^*) = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$

$L \subseteq \Sigma^*$

$L^* = \{\varepsilon\} \cup L \cup LL \cup LLL \cup \ldots$

Regexes are used extensively in text processing, text search, compilers (recognizing tokens)

Say that regex r <u>matches</u> string w (or w matches r) if $w \in L(r)$.

A ~~leag~~ legal identifier starts with a letter (or underscore) followed by 0 or more letters, ~~num~~ digits, or _

[A-Za-z_][A-Za-z_0-9]*

<u>Syntactic sugar</u> (shorthands) Σ = ~~ACS~~ ASCII char set

char classes

[acdf_] = a + c + d + f + _

[a-z] = [abcd ... z]

$\varepsilon := \emptyset^*$

5 $L(\varepsilon) = L(\emptyset^*)$

$= \{\varepsilon\} \cup \emptyset \cup \emptyset\emptyset \cup \emptyset\emptyset\emptyset \cup \cdots$

$= \{\varepsilon\}$

$r+$ (unary plus: one or more occurrences of r)

$r+ := rr^*$

$r? := r + \varepsilon$ (optional r: 0 or 1 occurrences of r)

. matches any single char

"." matches period

$r|s$ means $r+s$

Floating point constants (unsigned) in Pascal:

1 or more decimal digits
followed by a period
followed by one or more digits
optionally followed by exponent,

e.g., e-16

E+3

E3

(6)

$$([0-9]+)"."([0-9]+)([Ee][+-]?[0-9]+)?$$

$[0-9]+"."$ $(E|e)$

Monday

Regex $\rightleftarrows$ ε-NFA