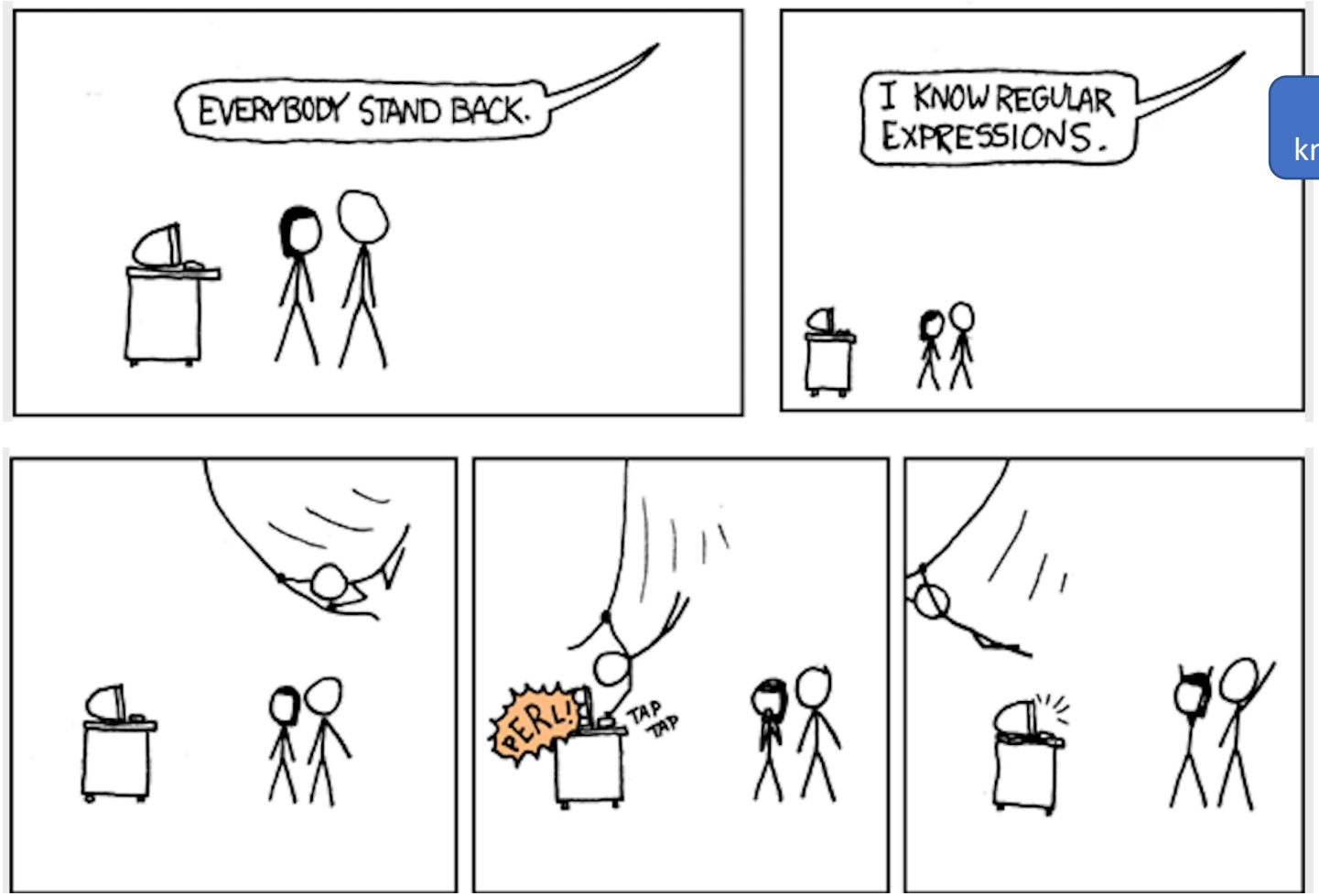
LEX3: Regexps are Trees

Lexical Analysis

CMPT 379: Compilers

Instructor: Anoop Sarkar

anoopsarkar.github.io/compilers-class



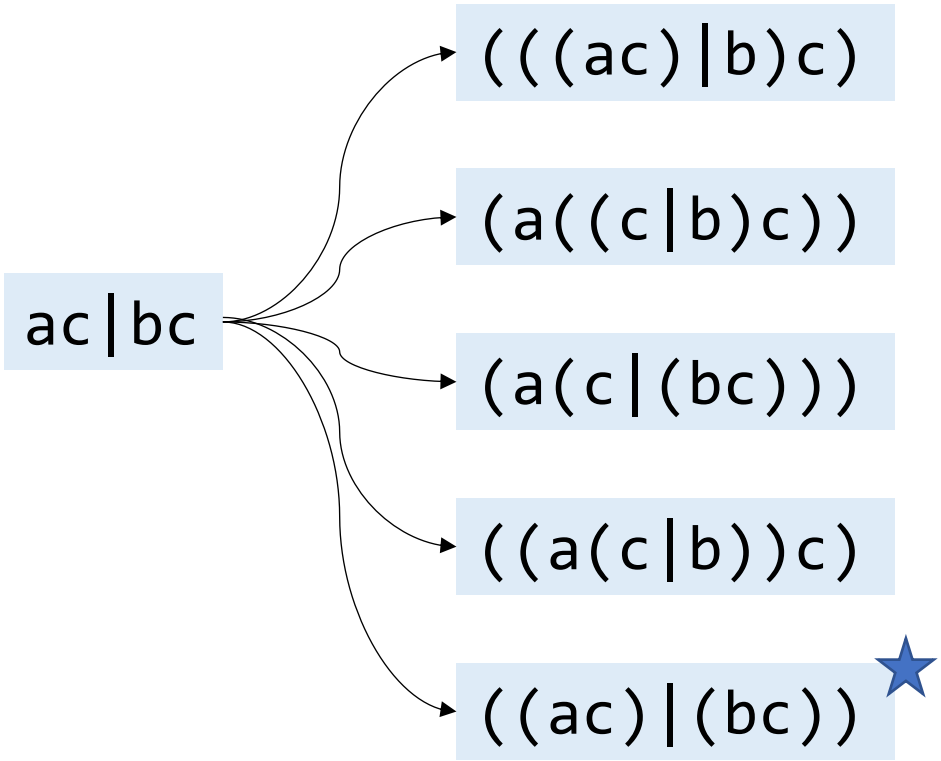
Nope! You don't know regexps (yet)!

xkcd.com/208

Regular Expressions are Trees

Regular Expressions are ambiguous

Q: How many possible trees for regexp ab*|c*



Regexp operator precedence rules

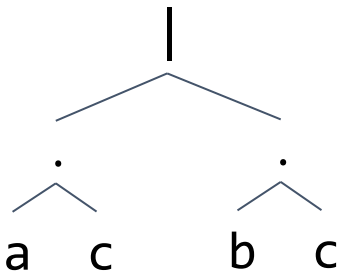
1. Grouping using parentheses ( )
2. Unary operator *
3. Binary operator for concatenation
4. Binary operator for alternation |

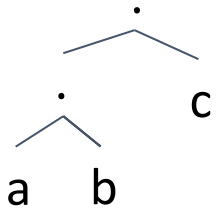
Q: Find the smallest set of strings that can find the above operator precedence rules for the regexp ac|bc.

Hint: Compare the matching on input strings between the original regexp ac|bc and the 5 unambiguous regexps.

Regular expressions are trees

Q: Provide the unambiguous bracketed tree for regexp ab*|c* using the precedence and associativity rules





Both alternation and concatenation are commutative, so we use left associativity to get a unique tree.

ac|bc

((ac)|(bc))

a**

((a*)*)

abc

((ab)c)

Equivalence of Regular Expressions

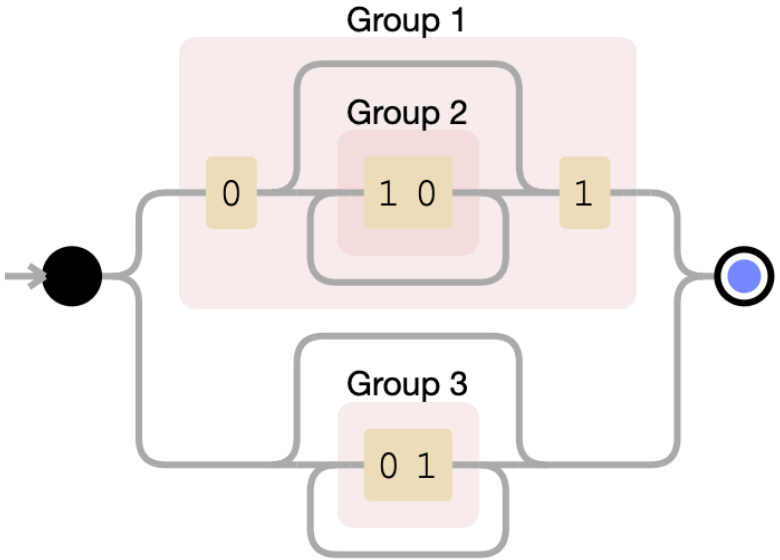
Equivalence of Regexps

(0(10)*1)|(01)* == (01)* ?

- (R|S)|T == R|(S|T)
- R|(S|T) == R|S|T
- (RS)T == R(ST)
- (R|S) == (S|R)
- R*R* == (R*)*
- (R*)* == R*
- R* == RR*| ε
- R** == R*
- (R|S)T = (RT|ST)
- R(S|T) == RS | RT
- (R|S)* == (R*S*)*
- (R*S*)* == (R*S)*R*
- (R*S)*R* == (R*|S*)*
- RR* == R*R
- (RS)*R == R(SR)*
- R == R|R
- R|R = Rε

Equivalence of Regexps

debuggex.com



(0(10)*1)|(01)* == (01)* ?

Equivalence of Regexps

Q: Show that `1(0|(01)*)` is equivalent to `10|(10)*1`

- (0(10)*1)|(01)*
- (01(01)*)|(01)*
- (01(01)*)|(01)*
- ((01)(01)*)|(01)*
- ((01)(01)*)|(01)*
- (01)+|(01)*
- (01)+|(01)*
- (01)*

(RS)*R == R(SR)*

RS == (RS)

R+ == RR*

R+|R* == (RR*)|R* == R*