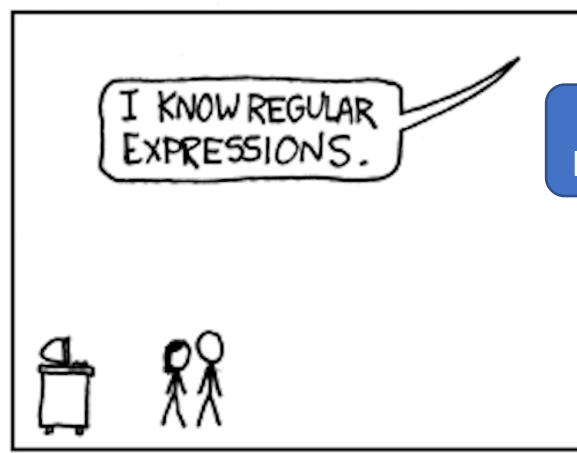


Lexical Analysis

CMPT 379: Compilers

Instructor: Anoop Sarkar

anoopsarkar.github.io/compilers-class



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

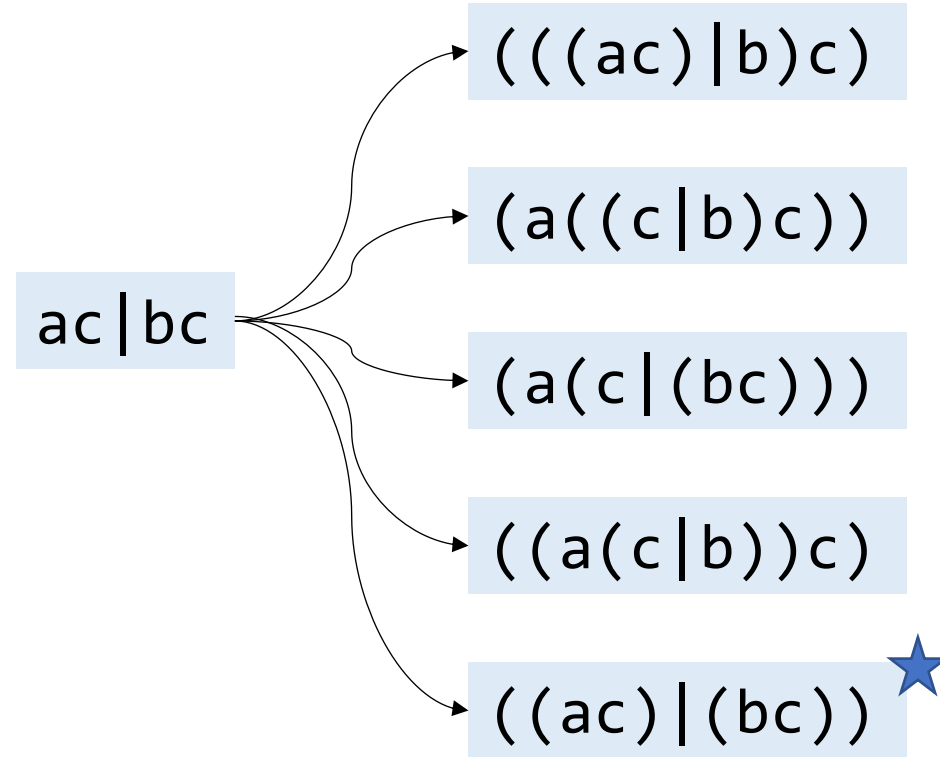


xkcd.com/208

Regular Expressions are Trees

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

Regular Expressions are ambiguous



Regexp operator precedence rules

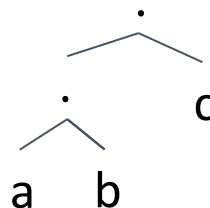
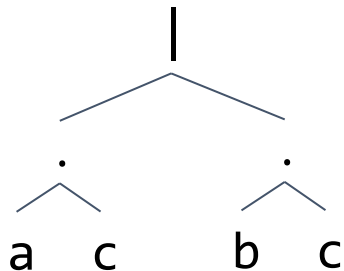
1. Grouping using parentheses ()
2. Unary operator *
3. Binary operator for concatenation
4. Binary operator for alternation |

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

a**

abc

$$((ac) | (bc))$$
$$(*(*a))$$
$$(. (.ab)c)$$

Equivalence of Regular Expressions

Equivalence of Regexp

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

- $(R|S)|T == R|(S|T)$
- $R|(S|T) == R|S|T$
- $(RS)T == R(ST)$
- $(R|S) == (S|R)$
- $R^*R^* == (R^*)^*$
- $(R^*)^* == R^*$
- $R^* == RR^* | \epsilon$
- $R^{**} == R^*$
- $(R|S)T = (RT|ST)$
- $R(S|T) == RS | RT$
- $(R|S)^* == (R^*S^*)^*$
- $(R^*S^*)^* == (R^*S)^*R^*$
- $(R^*S)^*R^* == (R^*|S^*)^*$
- $RR^* == R^*R$
- $(RS)^*R == R(SR)^*$
- $R == R|R$
- $R|R = R\epsilon$

Equivalence of Regexp

- $0(10)^*1|(01)^*$

$$(RS)^*R == R(SR)^*$$

- $01(01)^*|(01)^*$

- $01(01)^*|(01)^*$

$$RS == (RS)$$

- $((01)(01)^*)|(01)^*$

- $((01)(01)^*)|(01)^*$

$$R+ == RR^*$$

- $(01)^+|(01)^*$

- $(01)^+|(01)^*$

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

- $(01)^*$