# Lexical Analysis 

CMPT 379: Compilers<br>Instructor: Anoop Sarkar<br>anoopsarkar.github.io/compilers-class


xkcd.com/208

## Regular Expressions are Trees

## 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 $a c \mid b c$.

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


## Regular expressions are trees



## $a c \mid b c$


abc

## Equivalence of Regular Expressions

## Equivalence of Regexps $\left(0(10)^{*} 1\right) \mid(01)^{*}=(01)^{*}$ ?

| - (RS)T $==$ R(ST) <br> Commutative <br> - (R\|S)|T == R|(S|T) <br> - $(R \mid S)==(S \mid R)$ | - $R^{*}==R R^{*} \mid \varepsilon$ <br> - $\mathrm{R}^{*} \mathrm{R}^{*}==\left(\mathrm{R}^{*}\right)^{*}$ <br> - $\left(\mathrm{R}^{*}\right)^{*}=\mathrm{R}^{*}$ |
| :---: | :---: |
| - (R\|S)T == (RT|ST) <br> Factor <br> - $R(S \mid T)==R S \mid R T$ | - $R^{*}==R^{*} R$ <br> - $(R S)^{*} R==R(S R)^{*}$ |
| - $R==R \mid R$ <br> - $R \mid R==R \varepsilon$ | - $(R \mid S)^{*}==\left(R^{*} S^{*}\right)^{*}$ <br> - $\left(R^{*} S^{*}\right)^{*}==\left(R^{*} S\right)^{*} R^{*}$ <br> - $\left(R^{*} S\right)^{*} R^{*}==\left(R^{*} \mid S^{*}\right)^{*}$ |

## Equivalence of Regexps



$$
\left(0(10)^{*} 1\right) \mid(01)^{*}==(01)^{*} ?
$$

## Equivalence of Regexps

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

$$
(R S) * R==R(S R)^{*}
$$

- (01(01)*)|(01)*
- (01(01)*)|(01)*

$$
R S==(R S)
$$

- ((01)(01)*)|(01)*
- ((01)(01)*)|(01)*

$$
\mathrm{R}+==\mathrm{RR} *
$$

- (01)+|(01)*
- (01)+|(01)*

$$
\sqrt{R+\mid R^{*}}==\left(R R^{*}\right) \mid R^{*}=R^{*}
$$

- (01)*

