



# Expressões regulares: Uma introdução

Prof. Jesús P. Mena-Chalco  
jesus.mena@ufabc.edu.br

30/03/2016

# Apresentação

## ▼ Professor:

Jesús P. Mena-Chalco (CMCC)

## ▼ Formação:

- Engenheiro da Computação.
- Mestre e doutor em Ciência da Computação.  
Instituto de Matemática e Estatística da USP.

## ▼ Sala 517-A, torre 2, 5º Andar.

## ▼ Áreas de pesquisa:

- Cientometria/Bibliometria, Reconhecimento de padrões.

**Imagine:**

Você tem **UM** problema e descobre que precisa usar **expressões regulares...**

Parabéns!

Agora você tem **DOIS** problemas!

# Expressões regulares

- ▼ Regular expressions
- ▼ Regex
- ▼ regexp

São cadeias de texto especiais utilizadas para busca/identificação de “padrões” textuais.

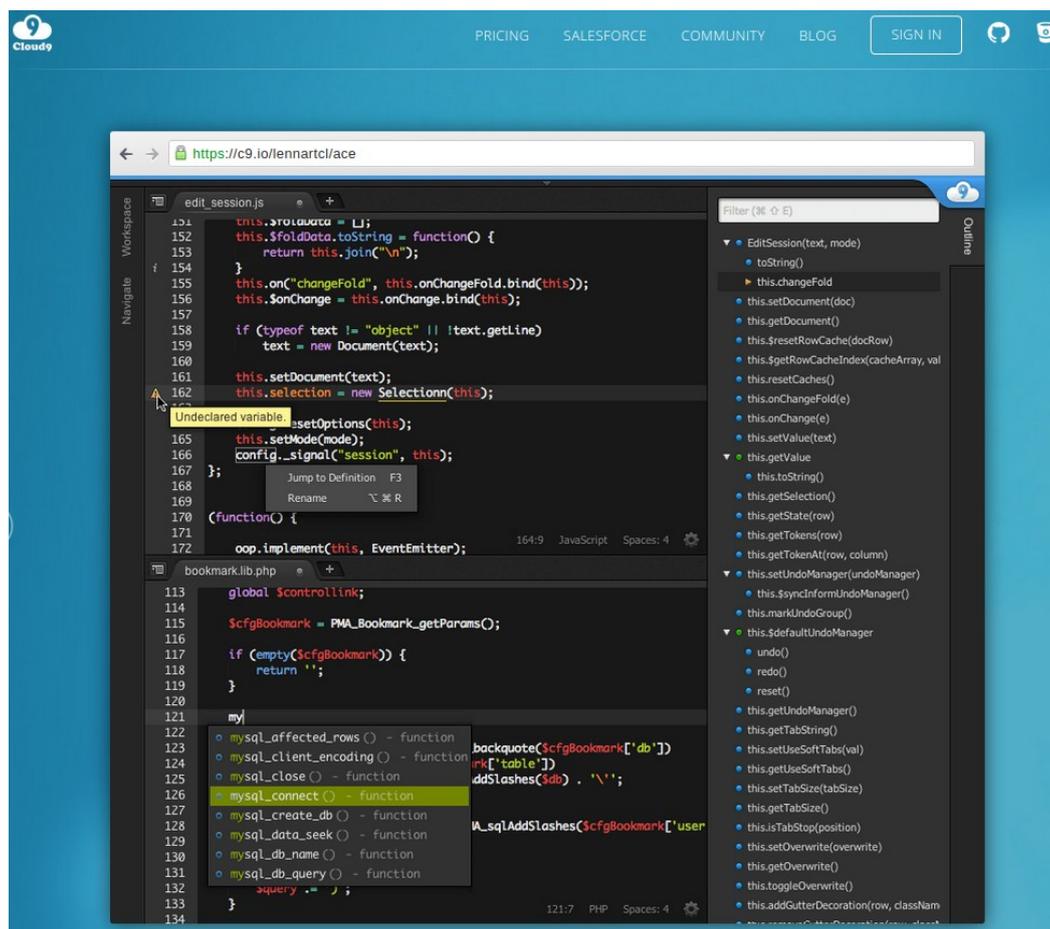
- ▼ \*.txt

- ▼ Maria\* → Maria, Mariana, Marianna, Mariano, Mariane

# Ferramenta

## ▼ Crie uma conta na plataforma Cloud9:

<https://c9.io/>



## Create a new workspace

Owner

J. P. Mena-Chalco

Workspace name

expressoes-regulares

Description

Make a short description of your workspace

[Hosted workspace](#)

[Remote SSH Workspace](#)



**Private**

This is a workspace for your eyes only



**Public**

This will create a workspace for everybody to see

Clone from Git or Mercurial URL (optional)

e.g. ajaxorg/ace or git@github.com:ajaxorg/ace.git

Choose a template



HTML5



Node.js



Meteor



PHP, Apache &...



Python

**django**

Django



Ruby



C++



Wordpress



Rails Tutorial



Blank

Create workspace

Cloud9 File Edit Find View Goto Run Tools Window Support Preview Run

MEMORY CPU Share

Workspace

expressoes-regulares

README.md

[M] /README.md

/README.md Markdown

# Cloud9 IDE

Hi there! Welcome to Cloud9 IDE!

To get you started, create some files, play with the terminal, or visit <http://docs.c9.io> for our documentation. If you want, you can also go watch some training videos at <http://www.youtube.com/user/c9ide>.

Happy coding! The Cloud9 IDE team

bash - "jmenac-expr... Immediate

```
jmenac:~/workspace $
```

Collaborate Outline Debugger

Cloud9 File Edit Find View Goto Run Tools Window Support Preview Run MEMORY CPU Share

Workspace  
expressoes-regulares  
README.md

[M] /README.md x +  
/README.md Markdown

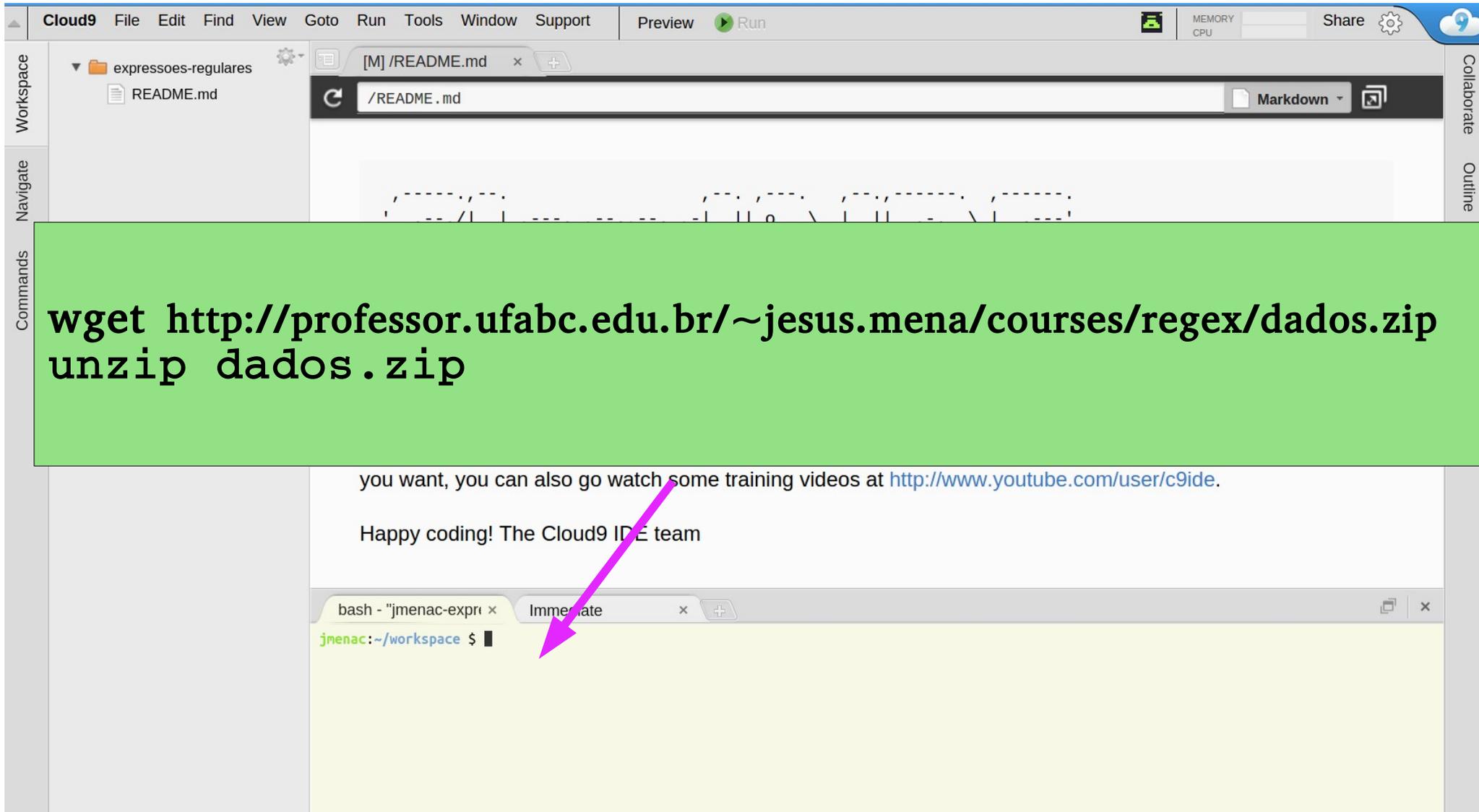
Collaborate Outline

```
wget http://professor.ufabc.edu.br/~jesus.mena/courses/regex/dados.zip
unzip dados.zip
```

you want, you can also go watch some training videos at <http://www.youtube.com/user/c9ide>.

Happy coding! The Cloud9 IDE team

bash - "jmenac-expr... x Immediate x +  
jmenac:~/workspace \$ █





# Localizando textos

# Expressões regulares

## ▼ Problema:

Quais disciplinas o “Paulo Ricardo” está cursando?

# Problema 1

- ▼ Quais disciplinas o “Paulo Ricardo” está cursando?



```
bash - "ubuntu@jme x"
jmenac:~/workspace $ egrep "Paulo Ricardo" alunos.txt
```

# Duas ferramentas: grep e egrep

- ▼ **grep (Global Regular Expressions Print):**

Programa que lê um arquivo (ou conjunto de arquivos), linha a linha, e imprime as linhas que contenham um padrão.

*Padrão:* é uma expressão que especifica um conjunto de cadeias interpretando caracteres como **meta-carateres**.

- ▼ **egrep (Extended grep)**

## ▼ Alguns parâmetros especiais

```
egrep -n "Paulo Ricardo" alunos.txt
```

```
egrep -i "Paulo Ricardo" alunos.txt
```

```
egrep -v "Paulo Ricardo" alunos.txt
```

# Problema 2

- ▼ Quais disciplinas o “Lucas Sousa” está cursando?

# Problema 2

- ▼ Quais disciplinas o “Lucas Sousa” está cursando?

```
egrep "Lucas .* Sousa" alunos.txt
```

# Problema 2

- ▼ Quais disciplinas o “Lucas Sousa” está cursando?

```
egrep "Lucas .* Sousa" alunos.txt
```

# Problema 2

- ▼ Quais disciplinas o “Lucas Sou**sa**” está cursando?
- ▼ Quais disciplinas o “Lucas Sou**za**” está cursando?

```
egrep "Lucas .* Sou[s|z]a" alunos.txt
```

# Problema 2

- ▼ Quais disciplinas o “Lucas Sou**sa**” está cursando?
- ▼ Quais disciplinas o “Lucas Sou**za**” está cursando?

```
egrep "Lucas .* Sou[s|z]a  " alunos.txt
```



2 espaços em branco

# Problema 3

- ▼ Quais os alunos que estão cursando a disciplina de Processamento da Informação C1, C2, C3, Diurno em Santo André?

```
egrep "Processamento da Informação [C1|C2|C3]-Diurno" alunos.txt
```



# Problema 3

- ▼ Quais os alunos que estão cursando a disciplina de Processamento da Informação C1, C2, C3, Diurno em Santo André?

```
egrep "Processamento da Informação [C1|C2|C3]-Diurno" alunos.txt
```

```
egrep "Processamento da Informação C[1|2|3]-Diurno" alunos.txt
```

```
egrep "Processamento da Informação (C1|C2|C3)-Diurno" alunos.txt
```



# Problema 3

- ▼ Quais os alunos que estão cursando a disciplina de Processamento da Informação C1, C2, C3, Diurno em Santo André?

```
egrep "Processamento da Informação (C1|C2|C3)-Diurno .*André" alunos.txt
```

```
egrep "Processamento da Informação (C1|C2|C3)-Diurno .*André" alunos.txt | wc
```

# Problema 4

- ▼ **Quais os alunos ingressantes em 2015?**

```
egrep "15" alunos.txt
```

# Problema 4

- ▼ Quais os alunos ingressantes em 2015?

```
egrep "15" alunos.txt
```

```
egrep "[0-9][0-9][0-9][0-9][0-9][0-9]15" alunos.txt
```

```
egrep "[0-9]{6}15" alunos.txt
```

# Problema 5

- ▼ Quantos os alunos ingressantes em 2015 e matriculados em 2016?
- ▼ Quantos os alunos ingressantes em 2006 e matriculados em 2016?

```
egrep -o "[0-9]{6}15" alunos.txt
```

```
egrep -o "[0-9]{6}15" alunos.txt | uniq | wc
```

```
egrep -o "[0-9]{6}06" alunos.txt | uniq | wc
```

```
egrep -o "[0-9]{6}15" alunos.txt | uniq > alunos2015.txt
```

# Problema 6

- ▼ **Quantos alunos estão matriculados em FUV e FVV?**

# Problema 6

- ▼ **Quantos alunos estão matriculados em FUV e FVV?**

```
egrep "Funções de Uma Variável" alunos.txt | egrep -o "[0-9]{8}" > alunosFUV.txt
```

```
egrep "Funções de Várias Variáveis" alunos.txt | egrep -o "[0-9]{8}" > alunosFVV.txt
```

```
cat alunosFUV.txt alunosFVV.txt | sort | uniq -d
```

```
cat alunosFUV.txt alunosFVV.txt | sort | uniq -d | wc -l
```

74

# Outros problemas

- ▼ No poema, “Agora José”, quantas vezes é feita a referência José?

```
egrep "José" jose.txt  
egrep -i "José" jose.txt  
egrep -i "José" jose.txt | wc -l
```

- ▼ Na letra da música, “Roda viva”, quantas vezes é feita a referência Roda?

Rodamoinho ão deve ser considerado:

# Outros problemas

- ▼ No poema, “Agora José”, quantas vezes é feita a referência José?

```
egrep "José" jose.txt  
egrep -i "José" jose.txt  
egrep -i "José" jose.txt | wc -l
```

- ▼ Na letra da música, “Roda viva”, quantas vezes é feita a referência Roda?

Rodamoinho ão deve ser considerado:

```
egrep -i "roda\b" rodaviva.txt
```



## **Localizando e substituindo textos**

expressoes-regulares

- alunos.txt
- jose.txt
- regexdados.zip
- rodaviva.txt

jose.txt

rodaviva.txt

```
36 sua gula e jejum,  
37 sua biblioteca,  
38 sua lavra de ouro,  
39 seu terno de vidro,  
40 sua incoerência,  
41 seu ódio - e agora?  
42  
43 Com a chave na mão  
44 quer abrir a porta,  
45 não existe porta;  
46 quer morrer no mar,  
47 mas o mar secou;  
48 quer ir para Minas,  
49 Minas não há mais.  
50 José, e agora?  
51  
52 Se você gritasse,  
53 se você gemesse,  
54 se você tocasse  
55 a valsa vienense,  
56 se você dormisse,  
57 se você cansasse,  
58 se você morresse...  
59 Mas você não morre,  
60 você é duro, JOSÉ!  
61  
62 Sozinho no escuro  
63 qual bicho-do-mato,  
64 sem teogonia,  
65 sem parede nua  
66 para se encostar,  
67 sem cavalo preto  
68 que fuja a galope,  
69 você marcha, José!  
70 José, para onde?
```

```
52 Se você gritasse,  
53 se você ficasse,  
54 se você ficasse  
55 a valsa vienense,  
56 se você ficasse,  
57 se você ficasse,  
58 se você ficasse...  
59 Mas você não morre,  
60 você é duro, JOSÉ!
```

bash - "ubuntu@jme

```
adding: alunos.txt (deflated 89%)  
adding: jose.txt (deflated 53%)  
adding: rodaviva.txt (deflated 68%)
```

jmenac:~/workspace \$

--  
52 Se você gritasse,  
53 se você gemesse,  
54 se você tocasse  
55 a valsa vienense,  
56 se você dormisse,  
57 se você cansasse,  
58 se você morresse...  
59 Mas você não morre,  
60 você é duro, JOSÉ!  
61  
62 Sozinho no escuro  
63 qual bicho-do-mato,  
64 sem teogonia,  
65 sem parede nua  
66 para se encostar,  
67 sem cavalo preto  
68 que fuja a galope,  
69 você marcha, José!  
70 José, para onde?

(15 Bytes) 53:16 Text Spaces: 4 

    se você.\*\b

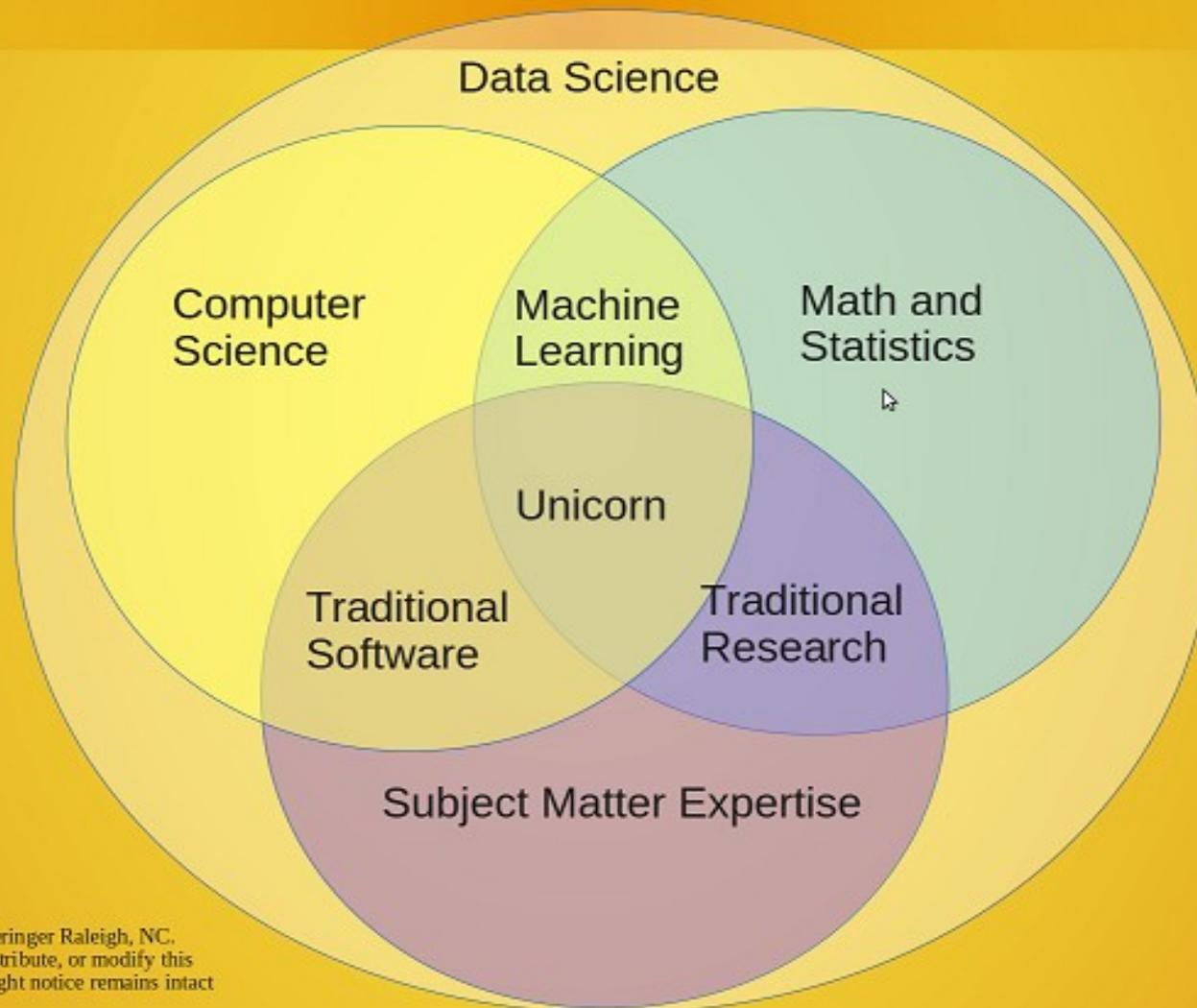
1 of 5 < >

 se você ficasse

Replace

Replace All 

# Data Science Venn Diagram v2.0



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this  
image, provided that this copyright notice remains intact