

Atividade: Extração de E-mails

Pedro Acácio Rodrigues

Universidade Tecnológica Federal do Paraná - Campo Mourão (UTFPR-CM)

Departamento de Computação - DACOM

pedrorodrigues.2019@alunos.utfpr.edu.br

Abstract. *The activity aims to perform the extraction of emails contained in a text file using regex.*

Resumo. *A atividade tem o objetivo de realizar a extração de emails contidos em um arquivo de texto utilizando regex.*

1. Problema

Escrever uma Expressão Regular em Python ou outras ferramentas que recupere do arquivo e-mails.txt a lista dos e-mails dos alunos da turma.

2. Base de dados utilizada

A base de dados utilizada contém os emails dos alunos da turma em um arquivo de texto (.txt)

```

"ARTUR MARCHI PACAGNAN" <marchiartur@gmail.com>,
"CAIO EDUARDO THEODORO DA SILVA" <caio-eduardotheodoro@hotmail.com>,
"ERIC FERREIRA RIBEIRO" <eribeiro.2000@alunos.utfpr.edu.br>,
"FABRICIO FLAVIO MARTINS DAMASCENO" <fabriciod@alunos.utfpr.edu.br>,
"GABRIELA MARANGONI RADIGONDA" <gabrielaradigonda@alunos.utfpr.edu.br>,
"GETULIO COIMBRA REGIS" <getulioaregis@alunos.utfpr.edu.br>,
"GREGULLY WILLIAN DA SILVA LIMA" <gregullywillian@alunos.utfpr.edu.br>,
"GUILHERME BERNARDO DEL RIO" <guilhermerio@alunos.utfpr.edu.br>,
"GUSTAVO SENGLING FAVARO" <gusfav@alunos.utfpr.edu.br>,
"HENRIQUE MOURA BINI" <henrique02@gmail.com>,
"IAGO SASAKI RAMOS" <iagosasaki@alunos.utfpr.edu.br>,
"IGOR LARA DE OLIVEIRA" <igooli@alunos.utfpr.edu.br>,
"JESSE PIRES BARBATO ROCHA" <jessepires2010@gmail.com>,
"JHONATAN GUILHERME DE OLIVEIRA CUNHA" <jhocunha1@gmail.com>,
"JOAO VICTOR NASCIMENTO" <nascimento783@gmail.com>,
"JULIANO KENDYI SHINOHARA" <julianoks1999@hotmail.com>,
"LUCAS ALEXANDRE SEEMUND" <seemund@alunos.utfpr.edu.br>,
"LUCAS HENRIQUE MALAQUIAS DA SILVA DONADI" <lucasdonadi@alunos.utfpr.edu.br>,
"MARCELO DANIEL GARCIA FILHO" <marcelodaniel@alunos.utfpr.edu.br>,
"MATHEUS HENRIQUE SECHINELI" <matheuss.2018@alunos.utfpr.edu.br>,
"MATHEUS KENJI NAKAO" <mnakao@alunos.utfpr.edu.br>,
"PEDRO ACACIO RODRIGUES" <pedrorodrigues.2019@alunos.utfpr.edu.br>,
"PEDRO HENRIQUE BORTOLUCCI KLAYN" <pedroklayn@alunos.utfpr.edu.br>,
"RYAN LAZARETTI RAMOS" <ryanramos@alunos.utfpr.edu.br>,
"VICTOR DANIEL MANFRINI PIRES" <victorpires.2000@alunos.utfpr.edu.br>,
"VINICIUS HENRIQUE SOARES" <viniciussoares@alunos.utfpr.edu.br>,
"VINICIUS KUWAKINO" <vikuwakino@gmail.com>,
"VINICIUS SIMOES VIEIRA" <vieira.1999@alunos.utfpr.edu.br>,
"Prof. Rogério" <rogerioag@professores.utfpr.edu.br>,
"Prof. Rogério" <rogerioag@utfpr.edu.br>,

```

Imagem 1. Base de dados

3. Execução

Para a extração de emails, a base de dados fornecida é submetida a expressão regular para localizar todos os emails contidos:

```
[\\w.]+@[a-z.]+\\. [a-z]{2,}
```

Na expressão regular utilizada são especificados possíveis emails que geram casos positivos na extração, dentre os possíveis podem conter “.” para separação, dois ou mais domínios no fim e a separação pelo “@”.

Assim a expressão regular analisa as ocorrências em emails, nele possuímos caracteres alfanuméricos podendo conter “.” no meio especificado pelo “\\w.”. O caractere “@” irá delimitar a separação entre o nome de usuário e o nome do domínio. “[a-z.]” simboliza uma ou mais ocorrências de caracteres alfabéticos minúsculos ou um ponto (.) após o símbolo “@” em um endereço de e-mail. “[a-z]{2,}” faz referência aos dois ou mais caracteres alfabéticos minúsculos que representam a extensão do domínio em um endereço de e-mail.’

4. Referencias

<https://docs.python.org/pt-br/dev/howto/regex.html>