



Projeto 1

Tarefa 1: Montar um dataset a partir de dados da web

18/03/15

Deadline

para submissão: 08/04/15

para comparação e avaliação dos resultados: 10/04/15

Entrega: Um script em R descrevendo os processos realizados e um arquivo .csv com o dataset.

Método de envio: Enviar para o repositório do GitHub (CCDrepository) ou para o email clubedecienciadedadosuff@gmail.com

Descrição

A primeira tarefa referente ao projeto será coletar todos os dados de ocorrências do site "onde fui roubado?" [1] com a linguagem R [2] através da técnica de Webscraping [3].

O dataset resultante deverá conter um formato padronizado de variáveis: nº da ocorrência, endereço, latitude, longitude, cidade, tipo de ocorrência, título da ocorrência, objetos roubados, data, hora e descrição da ocorrência.

Deverá ser entregue no dia 08/04/15 um Script em R descrevendo os processos realizados e um dataset no formato padronizado para facilitar a avaliação peer-to-peer.

Enviar para o repositório no GitHub (criar uma pasta dentro de /tarefas/projeto_1/tarefa_1 com o seu nome nele) ou por email (onde será postado no GitHub também para posterior avaliação)



Referências

- [1] <http://www.ondeluirobado.com.br/>
- [2] <http://www.r-project.org/>
- [3] http://en.wikipedia.org/wiki/Web_scraping



CCD-UFF
Clube de Ciência de Dados