

Fonte e Estrutura dos Dados

Os dados utilizados neste projeto são de domínio público e foram obtidos por meio do [Twitter API](#) por um script PHP e armazenados numa base MySQL, de onde foram exportados em formato .csv. Como os conjuntos de dados em formato .csv são muito pesados (entre 200 e 500 Mb), é conveniente salvar os dados em .rda, o que leva a uma redução bastante grande do tamanho do arquivo:

```
#carrega arquivo csv gerado na base MySQL e salva no formato. rda
hashdata <- read.table("filename.csv", header=TRUE, sep="|",
na.strings="NA", dec=".", strip.white=TRUE)
save("hashdata", file="freeiran.rda")
```

Estrutura dos Dados



text - mensagem do tweet com até 140 caracteres

from_user_id - identidade do usuário que postou a mensagem

iso_language_code - idioma do usuário

time - instante da publicação da mensagem (em [tempo Unix](#))

friends_count - número de amigos (Following/Friends)

followers_count - número de seguidores (Followers)

friends_ids - lista de IDs dos amigos

followers_count - lista de IDs dos seguidores

Os dados de friends e followers estão disponíveis como listas de IDs numéricos separados por vírgulas que em cada registro estão vinculados a um ID único do usuário que postou a mensagem



From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2011:alunos:trabalho_final:rafael:fonte



Last update: **2020/08/12 06:04**