



Reddit Viewer



Ana França (acff)
Édipo Andersen (eafs2)
Rodrigo Figueiredo (rff2)



Roteiro

- Motivação
- Dataset
- Possíveis análises
- Técnicas de visualização

Reddit

- Fórum de discussões
- Anonimato
- Conteúdo priorizado
- Espaço para debates com possibilidade de votação
 - i.e. um sistema de organização distribuída, e portanto extremamente escalável
- Subreddits (subfóruns)



Reddit

- 174 milhões visitantes únicos de 186 países diferentes *
- 6 bilhões de visualizações de página*
- Mais de 800 mil subreddits
 - Monitorados e geridos por moderadores voluntários
- Construção e debate de ativismo digital
- “O valor do usuário na comunidade é medido pelo conteúdo que ele compartilha, e não pelas pessoas com quem ele interage.”

*<https://www.redditsecrets.com/> (Agosto, 2016)

Datasets

- Api Oficial*
- Dataset minerado por Linan Qiu**
 - Licença *MIT*
 - Contém cerca de 260.000 tópicos/comentários minerados do Reddit (Abril, 2016).
 - 51 subreddits divididos em 7 metas: entertainment, gaming, humor, learning, lifestyle, news e television.



*<https://www.reddit.com/dev/api/>

** <https://github.com/linanqiu/reddit-dataset>

Informações do dataset

- Dataset minerado por Linan Qiu



anime comicbooks movies
starwars **ENTERTAINMENT** harrypotter music

adviceanimals
funny **HUMOR** jokes
circlejerk imgoingtohellforthis

askhistorians
science **LEARNING** askscience
youshouldknow space

dota2 minecraft
tf2 **GAMING** pokemon
starcraft skyrim

drunk food
sex **LIFESTYLE** motorcycles
guns lifehack

conspiracy politics
NEWS
libertarian conservative

community gameofthrones
TELEVISION
startrek doctorwho

Dataset

Atributo	Descrição
text	texto do tópico
id	identificador do tópico
subreddit	fórum ao qual o tópico pertence
meta	categoria a qual o tópico pertence
time	data da postagem
author	nome do usuário
ups	votos a favor
downs	votos contra
authorlinkkarma	upvote global de posts
authorkarma	upvote global de comentários
authorisgold	booleano indicando se o usuário é do tipo 'gold'



Linan Qiu

Análises - O que poderemos descobrir?

- Semânticas

- Que categorias são mais discutidas?
- Que categorias são mais acompanhadas?
- Com que frequência os usuários participam das discussões?
- Com que frequência um usuário se liga a vários subfóruns?
- Como os interesses dos usuários estão relacionados?
- Quais são os tópicos mais famosos (mais relevantes e bem discutidas)?

- Temporais

- Em que momentos as pessoas escrevem mais postagens?
- Como varia a intensidade das discussões nas categorias ao longo do tempo?

Técnicas de visualização

- Histogram
- Scatterplot
- Stacked Graph (TIARA)
- Force Directed Node

