

Artigo Tecnológico

Curtir, Compartilhar e Reagir: Captura de Tweets para Pesquisas e Decisões Corporativas

Like, Share, and React: Twitter Capture for Research and Corporate Decisions



Djeison Siedschlag*¹
Jeferson Lana²
Roberto Gonçalves Augusto Junior²
Rosilene Marcon¹

RESUMO

Contexto: as mídias sociais dispõem de uma imensa quantidade de informações, sendo um espaço para a sua difusão. Indivíduos, conexões on-line, são capazes de filtrar ou dar visibilidade a determinadas informações, em detrimento de outras. O problema central reside no monitoramento de postagens e reações voltadas a ações e estratégias corporativas. Além desse monitoramento, as empresas podem tomar decisões a partir dos dados coletados. **Objetivo:** desenvolver e estruturar uma ferramenta de gestão em mídia social. **Métodos:** para alcançar o objetivo geral, o artigo foi desenvolvido em três etapas principais. A primeira foi sugerir um script em software livre para captura e análise inicial das postagens no Twitter. A segunda etapa foi categorizar essa análise e identificar recursos e competências necessários às empresas. Por fim, foram sugeridas ações a serem tomadas pelas empresas para a gestão da mídia social. **Resultados:** o script desenvolvido possibilitou a extração dos dados de forma automatizada, os quais foram armazenados em banco de dados para análise e gerenciamento das interações on-line. As ações foram propostas com base no estudo de caso desenvolvido. **Conclusões:** no campo prático, este estudo contribuiu para o processo de extração de dados do Twitter com a proposição de um novo script para captura de dados, na identificação das principais categorias de influência de ativistas digitais e no monitoramento de mídia social por meio de ações estratégicas. Ao demonstrar que o script é eficaz na extração de dados, é possível realizar novos estudos e implementar o processo de monitoramento gerencial de mídia social.

Palavras-chave: ativismo social on-line; monitoramento de mídias sociais; extração de dados do Twitter; dados secundários.

ABSTRACT

Context: social media have an immense amount of information, being a space for its dissemination. Individuals, online connections, are able to filter or give visibility to certain information, to the detriment of others. The central problem lies in monitoring posts and reactions aimed at corporate actions and strategies. In addition to this monitoring, companies can make decisions based on the data collected. **Objective:** to develop and structure a social media management tool. **Methods:** to achieve the general objective, the article was developed in three main steps. The first was to suggest a free software script for capturing and initial analysis of Twitter posts. The second step was to categorize this analysis and identify resources and competencies needed by companies. Finally, actions to be taken by companies for social media management were suggested. **Results:** the developed script enabled the automated extraction of data, which were stored in a database for analysis and management of online interactions. The actions were proposed based on the case study developed. **Conclusions:** in the practical field, this study contributes to the process of extracting data from Twitter by proposing a new script for capturing data, identifying the main categories of influence of digital activists and monitoring social media through strategic actions. By demonstrating that the script is effective in extracting data, it is possible to carry out further studies and implement the social media management monitoring process.

Keywords: on-line social activism; social media monitoring; Twitter data extraction; secondary data.

* Autor Correspondente.

1. Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração, Biguaçu, SC, Brasil.
2. Universidade do Vale do Itajaí, Itajaí, SC, Brasil.

Como citar: Siedschlag, D., Lana, J., Augusto, R. G., Junior, & Marcon, R. (2023). Curtir, compartilhar e reagir: Captura de tweets para pesquisas e decisões corporativas. *Revista de Administração Contemporânea*, 27(2), e220008. <https://doi.org/10.1590/1982-7849rac2023220008.por>

Publicado em Early Access: 13 de Setembro, 2022.
Designado a essa edição: 11 de Janeiro, 2023.

Classificação JEL: M1.

Editor-chefe: Marcelo de Souza Bispo (Universidade Federal da Paraíba, PPGA, Brasil)
Editor Associado: Gustavo da Silva Motta (Universidade Federal Fluminense, PPGA, Brasil)

Pareceristas: Márcio Abdalla (Universidade Federal Fluminense, Brasil)

Dois revisores não autorizaram a divulgação de suas identidades.

Relatório de Revisão por Pares: A disponibilização do Relatório de Revisão por Pares não foi autorizada pelos revisores.

Recebido: 13/09/2021

Última versão recebida em: 30/05/2022

Aceite em: 11/06/2022

de revisores convidados até a decisão:

	1	2	3	4	5	6	7	8	9
1ª rodada	3	2	1	1					
2ª rodada	3	1	2	1					

INTRODUÇÃO – DA TEORIA À PRÁTICA

Monitorar é uma etapa fundamental na rotina de administrar. Exemplos da importância de monitoramento são plurais nas mais diversas áreas do saber: exames fisiológicos rotineiros ajudam a identificar alterações de saúde que podem gerar complicações mais graves; o acompanhamento climático permite antecipar desastres ambientais; assim como avaliações mensais permitem ao professor medir o grau de absorção do conteúdo lecionado. Ao passo que o planejamento direciona, a monitoração ajuda os gestores na correção da rota.

Com o desenvolvimento da tecnologia digital, a monitoração avançou a passos ainda mais acelerados. A potência computacional utilizada na captação e análise de dados permite às empresas avaliarem, em tempo real, as volatilidades das preferências de mercado. Embora existam discussões sobre como as empresas devem utilizar esses dados de forma estratégica, a captação dos dados e sua consequente análise tornaram-se indispensáveis, pois o uso generalizado da internet, de mídias sociais e sites gratuitos de compartilhamento de arquivos e vídeos facilitou a transmissão de informações persuasivas sobre questões diversas, bem como informações envolvendo organizações ou seus líderes em relação a problemas sociais (Briscoe & Gupta, 2016).

Diante da possibilidade de se manifestar a partir de alguns cliques, cidadãos, movimentos sociais, grupos diversos e até partidos políticos vêm usando esse tipo de plataforma na internet como ferramenta para disseminar diferentes tipos de informações. A finalidade dessas ações também é bastante diversa, podendo atender tanto a interesses democráticos como ao direito à informação e à liberdade de expressão, assim como à propagação de *fake news* e manipulação da opinião pública (Fundação Instituto de Administração [FIA], 2021).

Na área das pesquisas sobre o ativismo social, existem escolhas metodológicas e desafios. Em particular, as informações da internet podem ser utilizadas tanto para análise das interações sociais e organizacionais como para as respostas corporativas ao ativismo on-line. Nesse sentido, é provável identificar rastros eletrônicos que revelem vários impactos intencionais e não intencionais do ativismo contra organizações. Esses resíduos podem ser usados, por exemplo, para mapear o surgimento dinâmico de uma rede de ativistas internos (e externos) ao longo do tempo, ou para um evento ao longo da história (Briscoe & Gupta, 2016).

Dada a natureza multifacetada do ativismo em torno das organizações, os estudos abordaram o fenômeno por meio de diversas lentes teóricas, empíricas e epistemológicas. Muita pesquisa foi desenvolvida partindo da sociologia tradicional, que se expandiu de um foco no papel do

Estado para estudar protestos em campos organizacionais envolvendo empresas, universidades, hospitais e outros tipos de atores de mercado (Davis & Thompson, 1994; Etter & Albu, 2021; Rao et al., 2000; Van Dyke et al., 2004).

Há um consenso na literatura sobre o fato de que a forma como os ativistas operam e influenciam as organizações está evoluindo ao longo dos anos com as tecnologias de informação e comunicação (TICs). Earl e Kimport (2011) argumentam que os avanços nas TICs aumentaram dramaticamente instâncias de ativismo on-line, reduzindo os custos de participação de ativistas e a necessidade de os movimentos sociais mobilizarem infraestrutura. Nesta linha, o estudo de Zhang e Luo (2013) mostrou que campanhas ativistas on-line usando plataformas de mídia social populares podem gerar concessões significativas e aumentar a filantropia corporativa.

A importância das mídias sociais para os movimentos de protesto contemporâneos foi teorizada a partir de diferentes perspectivas. No geral, enfatizou-se a funcionalidade das mídias sociais como meio de compartilhamento de informações e ferramenta para organizar protestos (Bimber, Flanagin, & Stohl, 2012). Menos atenção tem sido dada ao potencial impacto de internautas que utilizam as mídias sociais para efetuar mudanças sociais e como as empresas responderão a essas demandas. Artefatos digitais ligados a protestos e outras formas de ativismo on-line oferecem valiosas informações, mas estas raramente são analisadas por pesquisadores (Jenzen et al., 2021).

No campo das pesquisas organizacionais, tais informações podem ser analisadas, por exemplo, por meio da mineração de dados (*data mining*) e aplicação de métodos estatísticos ou análises qualitativas (Mercado & Silva, 2013; Ross & Cruz, 2021). No campo prático, todo o volume de dados gerado nas mídias sociais precisa ser analisado, em um processo de acompanhamento e mensuração cuidadoso e focado. Isso é feito por meio do monitoramento das mídias sociais, em que as empresas acompanham como o público interage e reage a determinados temas em seus canais digitais.

Se, por um lado, o ativismo on-line é um fenômeno de relevância nos meios acadêmico e organizacional, por outro lado, mudanças drásticas a partir do ano de 2018 marcaram de forma negativa a história das mídias sociais e suas interfaces de programação de aplicações ou APIs (sigla para *application programming interface*, em inglês), que se referem ao conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades (Meirelles, 2019). Ou seja, se antes pesquisadores conseguiam coletar e monitorar todas as publicações feitas no mural/*feed* dos usuários, desde que estivessem em modo público (de exposição), a partir da descontinuidade e restrições das APIs tal prática não foi mais possível, o que impactou, de

forma significativa, o processo de monitoramento das mídias sociais. A partir dessa problemática, o objetivo central deste estudo é desenvolver e estruturar uma ferramenta de gestão em mídia social.

ASPECTOS CONCEITUAIS

O termo 'mídia social' refere-se à prática de usar qualquer meio específico socialmente. Também é comumente utilizado para descrever aplicativos da web, como sites e plataformas de compartilhamento de conteúdo gerado por usuários (Kaplan & Haenlein, 2010). Os autores destacam que 'mídia social' se refere a plataformas dinâmicas de internet e à sua capacidade de apoiar a interatividade e a comunicação em tempo real.

A adoção das mídias sociais pela sociedade tem levado as organizações a adotarem ferramentas tecnológicas para divulgar informações organizacionais importantes para seus *stakeholders* (Kim & Youm, 2017; Smits & Mogos, 2013). Conforme Lee, Hutton e Shu (2015), as mídias sociais se tornaram um canal viável para divulgação de informações estratégicas, quando comparadas aos canais tradicionais de informação. Elas possibilitam um maior alcance e permitem que as empresas enviem de forma direta e com rapidez a mensagem pretendida.

Contudo, a mídia social pode ser vista sob duas perspectivas antagônicas. A primeira, de uma forma positiva, caracteriza-se por diminuir o custo de comunicação, ampliar a velocidade da informação e eliminar dificuldades em relação à distância física. Na segunda, negativa, destaca-se o empobrecimento do debate político, com discussões mais vazias (predominando a simplificação), o uso de notícias falsas como estratégia para reiterar ideologias e a perda de privacidade no ciberespaço (Cavalcanti et al., 2019).

Inicialmente, as mídias sociais foram utilizadas pelos gestores para fins de marketing (Arnaboldi et al., 2017). Posteriormente, organizações passaram a usar essas ferramentas principalmente para transmitir a sua imagem idealizada (Harquail, 2011). Esse meio de divulgação vem ganhando popularidade e sendo cada vez mais utilizado em operações regulares, por muitas empresas (Smits & Mogos, 2013).

Nesta dinâmica, aqueles que utilizam as tecnologias mais recentes das mídias sociais parecem superar seus concorrentes e relatam alguns benefícios, como custos mais baixos e melhor eficiência (Harris & Rea, 2009). Portanto, pesquisadores relatam a importância de entender a relação entre o uso das mídias sociais e o desempenho de organizações (Smits & Mogos, 2013; Ghardallou, 2021; Wetzstein et al., 2011).

Comparada com a mídia tradicional, a internet é muito mais difícil de regular. Além disso, as mídias sociais e a comunicação digital ampliam significativamente a participação de diversos atores em ação coletiva. Nesse contexto, as mídias sociais têm sido destacadas por pesquisadores e gestores, em especial pela sua pressão por respostas rápidas e o seu impacto no desempenho de organizações (Cavalcanti et al., 2019; Ghardallou, 2021; George & Leidner, 2019; Gomez-Carrasco & Michelon, 2017; Lewis, 2005; Luo et al., 2016; Van Dijck & Poell, 2013).

Dada a relevância das mídias sociais para as pesquisas no campo organizacional, torna-se crucial o seu entendimento. Evidências anedóticas demonstram a importância das mídias sociais. Em 2021, a base do aplicativo Twitter no Brasil, por exemplo, foi de aproximadamente 17,46 milhões de usuários. O número de usuários no Brasil está projetado para atingir 18,39 milhões até 2025 (Statista, 2021).

No campo acadêmico/prático, o Twitter tem significativa relevância para o entendimento das ações e reações de pessoas sobre determinado tópico ou mesmo para mensurar a reputação e estratégia organizacional. Em recente pesquisa, Abdulaziz, Alotaibi, Alsolamy e Alabbas (2021) estudaram, por meio do Twitter, os principais tópicos e mudanças nas preocupações das pessoas acerca da pandemia de COVID-19.

Na pesquisa de Shin e Ki (2022), os autores exploraram, por meio dos *tweets*, como organizações com e sem fins lucrativos se posicionam em relação ao meio ambiente e à resposta do público. A análise mostrou que empresas com fins lucrativos tendem a discutir seus produtos verdes, enquanto as organizações sem fins lucrativos estão dispostas a descrever um ambiente degradado. Além disso, o estudo revelou que os *tweets* geram um alto número de curtidas e respostas quando as organizações têm fins lucrativos e as mensagens enfatizam produtos verdes.

Vogler (2020) avaliou a reputação de universidades suíças utilizando o Twitter. Segundo o autor, as reações aos *tweets* na forma de *retweets* e curtidas indicam quanta atenção os *tweets* receberam. As reações ao conteúdo das mídias sociais dependem da forma como o conteúdo é apresentado, sendo as emoções um importante preditor da chamada viralidade.

Independentemente da temática de pesquisa, o Twitter é uma importante fonte de informações para estudos organizacionais. Para Jung, Naughton, Tahoun e Wang (2018), o Twitter corresponde a uma das plataformas de mídia social mais usadas para fins corporativos. Porém, para a coleta de dados, a atual API limita o número de *tweets* coletados, o que impossibilita estudos que demandem dados históricos de organizações, por exemplo. Adicionalmente,

não está claro, na literatura atual, quais ações gerenciais em nível estratégico as empresas devem desenvolver para o monitoramento e gerenciamento de mídia social, sendo este o problema central a ser discutido neste estudo.

Application programming interfaces

Para auxiliar no processo de integração entre sistemas, existem as *application programming interfaces* (APIs) (ou interfaces de programação de aplicações, em português). As APIs são interfaces que funcionam como pontes, transportando dados entre um cliente e um servidor.

Sem que esse processo seja sequer percebido pelo usuário, as APIs estão presentes no funcionamento de diversos programas e aplicativos. Elas são séries de comandos que permitem a usuários e aplicativos se comunicarem com os sites e requisitarem dados hospedados em seus servidores (Silva & Stabile, 2016).

Seu funcionamento dá-se, em geral, quando o cliente precisa ter acesso a determinados dados disponibilizados por um banco de informações, serviço ou dispositivo, por exemplo. Para acessar essas informações específicas, o cliente faz uma solicitação – nomeada de requisição – à API. Assim, a interface consegue fazer uma busca no servidor e devolver as respostas solicitadas em formato de dados, que são entregues em seu estado mais puro.

D'Andréa (2021) destaca que, por meio das APIs, plataformas on-line oferecem acesso a uma parte dos dados por elas gerados e/ou coletados, como os gêneros musicais atribuídos a um artista (Spotify), as postagens vinculadas a um termo ou a uma *hashtag* (Twitter) ou o total de visualizações, curtidas e comentários, até um dado momento, dos vídeos postados por um canal (YouTube).

No que tange à segurança, por meio de uma API, é possível ter maior controle sobre as permissões de acesso ao software e hardware de uma empresa. Ela identifica quem são as pessoas que tentaram acessar o sistema e localiza de onde partiu a requisição. Além disso, é possível configurar quais informações ficarão disponíveis ao cliente no momento da integração.

Segundo Meirelles (2019), o ano de 2018 marcou – de forma negativa – a história das APIs das mídias sociais. Depois de tantos escândalos (políticos) envolvendo o uso de dados de maneira irresponsável, o conglomerado do Facebook se viu na contingência de agir publicamente na tentativa de argumentar que – pelo menos a partir daquele momento – estaria mais comprometido com a privacidade e os dados de seus usuários. Segundo o autor, esse fato impactou diretamente o mercado de monitoramento de mídias sociais, que viu sua fonte de dados diminuir ainda mais.

Sob um olhar mais abrangente, pode-se afirmar que qualquer pessoa – pesquisador(a) independente ou acadêmico(a) – que trabalhou com dados de mídias sociais acabou sofrendo com as últimas decisões da empresa. Essa realidade, entretanto, já não é novidade: nos últimos anos, mudanças drásticas na API já apontavam para restrições mais severas (Meirelles, 2019).

Conforme a linha de tempo (Figura 1), as APIs do Facebook e do Twitter foram lançadas oficialmente em 2006. Em abril de 2010, o Facebook lançou a primeira versão da Graph API.

As primeiras ferramentas de monitoramento foram adotadas pelo mercado brasileiro ao final desse período, com nomes como Scup, Radian6, Trendr e Socialmention.

Conforme Meirelles (2019), nesse primeiro momento, a quantidade infinita de dados disponibilizados pelo Facebook e Twitter já era vista com muito otimismo por diferentes atores: as próprias empresas, os profissionais de marketing, pesquisadores (acadêmicos), etc.

No entanto, após anos de muito crescimento e entusiasmo, começaram a surgir os primeiros problemas, principalmente para o Facebook. Já entre 2010 e 2013, portanto, algumas matérias ‘denunciavam’ a quantidade absurda de dados e informações sobre seus usuários que o Facebook mantinha – e que compartilhava facilmente com terceiros.

Depois de quatro anos de acesso praticamente ilimitado à infinidade de dados dos usuários, a v1.0 da Graph API do Facebook foi enfim descontinuada em abril de 2014. Ou seja, se antes as ferramentas conseguiam coletar todas as publicações feitas no mural/*feed* dos usuários, desde que estivessem em modo público (de exposição), a partir da descontinuidade da API isso já não seria mais possível.

Depois do baque de 2015, o mercado de monitoramento e pesquisa em mídias sociais foi obrigado a reformular o modo como, na prática, trabalhava com os dados. A maior mudança nesse sentido foi o redirecionamento de atenção aos dados do Facebook: uma vez que já não era mais possível acompanhar o que as pessoas compartilhavam em seus perfis públicos, as ferramentas – e o mercado – voltaram-se principalmente para outras mídias sociais.

No final de 2017, o Twitter anunciou o lançamento de APIs *premium*. A proposta era encontrar um meio-termo entre os limites da API gratuita (e pública), mais utilizada pelas ferramentas de monitoramento e pesquisadores independentes, e o turbilhão de dados da API *enterprise*, o chamado *firehose* disponível pelo GNIP (que dava acesso a todos os *tweets* da plataforma desde 2006). Segundo o Instituto Brasileiro de Pesquisas e Análise de Dados (IBPAD) (2021), o Twitter sempre manteve uma postura



Figura 1. Linha do tempo das APIs.

Fonte: Adaptado de Meirelles, P. (2019). Histórico das APIs no monitoramento e pesquisa em mídias sociais. Brasília, DF: Instituto Brasileiro de Pesquisa e Análise de Dados.

favorável à requisição de dados via API. Desenvolveu, ao longo dos anos, diferentes ‘opções’ de APIs, para atender principalmente à demanda do mercado: enquanto, de um lado, a versão gratuita sempre foi de fácil acesso (com limite de sete dias retroativos e requisição a cada 15 minutos), na outra ponta, a já denominada *firehose* (ou *enterprise*) custava um preço muito alto para empresas interessadas no acesso ilimitado às publicações da plataforma (com retroatividade até 2006, quando foi fundada). Para preencher essa lacuna (com possibilidade retroativa de até um mês e maior volume de dados na requisição), a empresa lançou as APIs *premium*.

Assim, a limitação das plataformas no acesso aos dados requer, de pesquisadores e gestores, formas alternativas que realizem o processamento, sistematização e gestão desses dados. Não basta apenas realizar a extração dos dados. É também necessário desenvolver uma análise capaz de auxiliar as empresas na tomada de decisão estratégica.

ANÁLISE DA SITUAÇÃO/PROBLEMA E SOLUÇÃO PROPOSTA

Nesta seção são apresentadas as etapas desenvolvidas para extração de dados da web, o script (conjunto de

comandos), a segmentação do processamento para a extração dos *posts*, as ações organizacionais e o fluxograma para resposta corporativa ao ativismo on-line.

Extração de dados da web

O tema proposto para realização deste estudo faz parte de uma pesquisa acadêmica, em que se analisaram os *tweets* das cem maiores instituições de ensino superior (IESs) no Brasil em número de alunos. Para tanto, foi observado se as IESs utilizavam mídia social e, então, buscou-se avaliar as postagens dos seus usuários no Twitter no período de 2013 a 2019. Devido à limitação de acessos a dados históricos da plataforma, o que inviabilizaria a análise retroativa das postagens realizadas, optou-se por desenvolver um script (conjunto de comandos) capaz de gravar dados de páginas web ao se acessar o site.

Para o desenvolvimento deste estudo, foi utilizado o algoritmo de *bot* da web e a técnica de *web scraping* para extração dos dados do Twitter. O *bot* da web é um algoritmo usado para analisar e extrair informações dos websites de forma sistemática e automatizada (Omari et al., 2016). Esses *bots* capturam informações das páginas web e cadastram

os links identificados para que possam ser posteriormente utilizados. Os *bots* fazem rastreamento e raspagem de dados utilizando duas técnicas: *web crawling* e *web scraping*. Essas técnicas podem ser usadas de forma simultânea ou como duas tarefas distintas (Khalil & Fakir, 2017).

Web crawling é o processo responsável por efetuar as buscas das páginas web e indexá-las. Essa técnica captura informações dos websites e cadastra os links encontrados, de forma que possam ser localizados futuramente (D’Haen et al., 2016). O processo é iniciado a partir de uma *seed* (semente), que consiste em um ponto de partida para encontrar novos endereços a serem visitados. À medida que o *bot* visita esses endereços, os hyperlinks são identificados e adicionados à lista de endereços a serem visitados (Khalil & Fakir, 2017), proporcionando a localização de outras páginas, o que mantém o banco de dados atualizado.

Web scraping se refere ao processo de extração utilizado para coletar dados diversos de sites de forma automática, convertendo as informações desestruturadas em estruturadas, para serem posteriormente analisadas (Zhao, 2017), em um procedimento conhecido como raspagem de dados. O *bot*, que utiliza a técnica *web scraping*, é programado para efetuar requisições a um servidor web a partir de uma

lista predefinida de URLs. Após a solicitação, são extraídos os dados necessários. Os dados obtidos são copiados e podem ser exportados em arquivos nos formatos JSON (sigla para notação de objetos JavaScript ou *JavaScript object notation*, em inglês), e valores separados por vírgulas (CSV, ou *comma separated value*, em inglês), entre outros. Normalmente, esse processo simula uma navegação humana na utilização de um site, porém, o *bot* consegue efetuar mais requisições do que as feitas por uma pessoa.

Cabe destacar que a mídia social foco deste estudo, o Twitter, adotou uma política de privacidade (Twitter, 2022a) que torna públicos todos os *tweets* registrados pelos seus usuários, ou seja, eles se tornam visíveis e pesquisáveis por qualquer pessoa. Conforme a política de privacidade do Twitter, ao compartilhar publicamente um conteúdo através de um *tweet*, o usuário concorda em divulgar essas informações da maneira mais ampla possível. Assim, o usuário é responsável por seus *tweets* e por outras informações fornecidas por meio dos serviços disponibilizados.

A construção do script deste artigo seguiu as diretrizes do Twitter, conforme pode ser consultado pelo link <https://twitter.com/robots.txt>. Na Figura 2 são apresentados os parâmetros utilizados no desenvolvimento do *bot*.

```
# Every bot that might possibly read and respect this file
# =====
User-agent: *
Allow: /*?lang=
Allow: /hashtag/*?src=
Allow: /search?q=%23
Allow: /i/api/
Disallow: /search/realtime
Disallow: /search/users
Disallow: /search/*/grid

Disallow: /*?
Disallow: /*/followers
Disallow: /*/following

Disallow: /account/deactivated
Disallow: /settings/deactivated

Disallow: /oauth
Disallow: /1/oauth

Disallow: /i/streams
Disallow: /i/hello
```

Figura 2. Parâmetros utilizados para desenvolvimento do *bot* Twitter.

Fonte: Twitter (2022b). Robots. Retrieved from <https://twitter.com/robots.txt>

A ferramenta foi desenvolvida seguindo as diretrizes da plataforma, bem como as boas práticas no desenvolvimento de software de acordo com o que estabelece a Lei n.º 13.709 (LGPD) (2018), que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade

e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

Script (conjunto de comandos) para extração de dados

Na tarefa de extração de dados do Twitter, o caminho mais objetivo é utilizar as ferramentas fornecidas pela

própria rede social. O Twitter, conforme citado, oferece uma API de extração de dados (Meirelles, 2019). Entretanto, esta API limita o usuário a extrair *tweets* recentes, publicados no máximo há sete dias da data de extração dos *posts*, inviabilizando, dessa forma, a extração de dados publicados anteriormente a esse período.

Publicações históricas são relevantes a uma quantidade significativa de análises, inclusive a leitura informal da rede social pelos usuários, que tendem a procurar postagens antigas. Além disso, o estabelecimento de uma estratégia

de monitoramento histórico de redes sociais pode trazer uma série de benefícios para a empresa como, por exemplo, a compreensão do comportamento do público através da observação do que lhe agrada ou não, além do entendimento da percepção da marca junto ao público que faz uso dessas mídias. A relevância dos dados históricos, além dos fornecidos pela API, é significativa, e o Twitter oferece, aos seus usuários, uma busca avançada, que lhes permite filtrar palavras-chave e intervalo de datas, conforme apresentado na Figura 3.

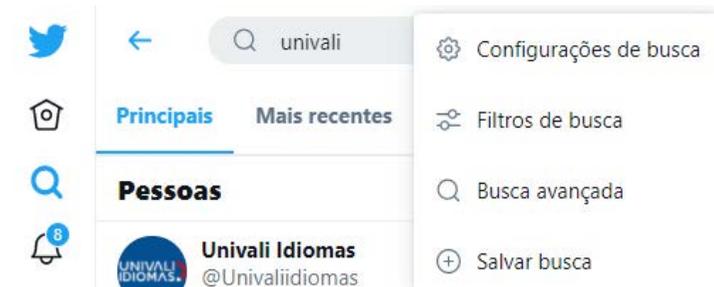


Figura 3. Busca avançada do Twitter.
Fonte: Dados da pesquisa – Twitter.

Ao necessitarem de uma análise sistemática das publicações históricas, apoiada em métodos estatísticos e ferramentas computacionais, os interessados na análise devem realizar a coleta e catalogações manualmente, para que o dado possa efetivamente ser analisado, o que demanda esforço significativo de pesquisadores e gestores. É nesta primeira etapa de coleta das publicações históricas que este trabalho apresenta relevância, pois a nova forma de consulta supera a limitação da API oferecida pela rede social, e reduz significativamente o tempo de coleta dos dados.

A ferramenta desenvolvida consiste em um script (conjunto de comandos) na linguagem de programação Python, o navegador Selenium e a biblioteca BeautifulSoup, para implementar um robô que simula a utilização da rede social por um humano, realizando as pesquisas manualmente, conforme esquema apresentado na Figura 4.

O script foi desenvolvido a partir do trabalho de Choudhary (2021), disponibilizado no site github.com. À época em que este artigo foi escrito, a última atualização daquela ferramenta ocorreu em 03/10/2017, não estando apta a extrair dados da versão atual do site do Twitter.

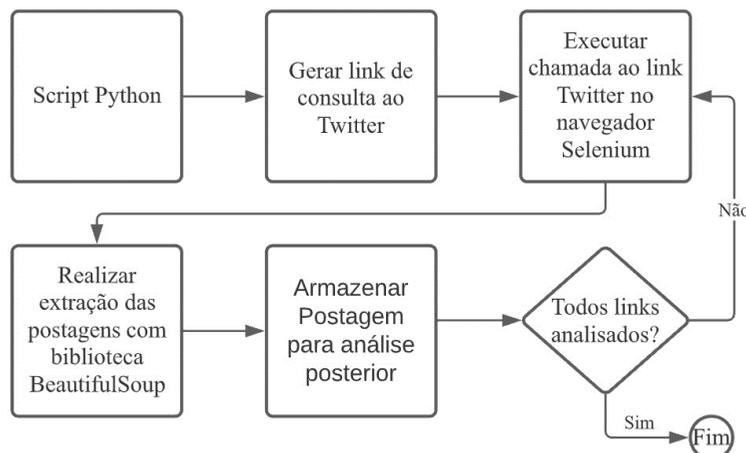


Figura 4. Esquema do robô para extração de dados.
Fonte: Elaborado pelos autores.

Na primeira etapa, utilizou-se a técnica de *web crawling* para rastrear os links das pesquisas geradas pela busca avançada, oferecida pela mídia social do Twitter. Para isso, utilizou-se um conjunto de palavras-chave para teste contendo como referência os termos que identificam as IESs em estudo. Cabe destacar, conforme descrito na introdução, que o tema proposto para a testagem do robô faz parte de uma pesquisa acadêmica, em que se analisaram os *tweets* de IESs no Brasil; todavia, poderiam ser utilizadas outras palavras-chave, de acordo com o objetivo de pesquisa.

Cada palavra-chave foi cruzada com cada um dos dias, no intervalo de 01/01/2013 a 31/12/2019, gerando assim uma tabela com três colunas, uma para a palavra-chave a ser consultada no Twitter, outra para as datas de referência, e a seguinte para as datas de referência + 1 dia, conforme

ilustrado na Tabela 1, tomando como exemplo a instituição de ensino superior Universidade do Vale do Itajaí (Univali).

Na busca avançada no Twitter, é possível informar um intervalo maior de datas (acima de um dia). Entretanto, optou-se por analisar *posts* dia a dia, para minimizar a possibilidade de um volume grande de dados em uma mesma página.

Para cada linha da Tabela 1, foi criado um link, que pode ser generalizado pelo seguinte modelo: <https://twitter.com/search?q=[PALAVRA_CHAVE]%20until%3A[DATA_REFERENCIA+1]%20since%3A[DATA_REFERENCIA]&src=typed_query>

Utilizando a primeira linha da Tabela 1, fizeram-se as seguintes substituições, conforme observado na Tabela 2.

Tabela 1. Dados para gerar consultas no Twitter.

Palavra-chave	Data de referência	Data de referência + 1
univali	01/01/2013	02/01/2013
univali	02/01/2013	03/01/2013
univali	03/01/2013	04/01/2013
...		
univali	31/12/2019	01/01/2020

Nota. O intervalo de datas utilizado nesta consulta é apenas ilustrativo, podendo ser utilizado qualquer intervalo desejado.

Tabela 2. Palavra-chave e data de referência.

Itens	Substituído por
[PALAVRA_CHAVE]	Substituído por 'univali' – exemplo
[DATA_REFERENCIA]	Substituído por 2013-01-01, devido ao formato ano-mês-dia, exigido pelo Twitter
[DATA_REFERENCIA]	Substituído por 2013-01-02, devido ao formato ano-mês-dia, exigido pelo Twitter

Nota. A palavra-chave 'univali', utilizada nesta consulta, refere-se ao objeto de pesquisa, podendo ser utilizado qualquer termo, conforme a necessidade de cada estudo.

Após a substituição dos dados das Tabelas 1 e 2 no link-modelo, foi gerado o seguinte link:

<https://twitter.com/search?q=univali%20until%3A2013-01-02%20since%3A2013-01-01src=typed_query>.

Este link retornou os *posts* em que a palavra-chave 'univali' foi citada em postagens do dia 01/01/2013. Não havendo postagens que atendessem a essa consulta, uma página sem postagens foi retornada, normalmente com alguma publicidade, assim como no exemplo da Figura 5, apresentada no navegador, ao ser digitado o link na barra de endereços.

No total, foram criados 265.811 links, que atendiam a consultas de 91 palavras-chave, multiplicadas por 2.921 dias.

Na segunda etapa, o robô executou a chamada ao link e aguardou que a página de resultados fosse disponibilizada pelos servidores da mídia social, simulando uma consulta realizada por um usuário humano. Neste acesso, foi utilizada a ferramenta Selenium, que nada mais é do que um navegador web, que pode realizar tarefas como clicar em um link, utilizando um script Python (Selenium, 2021).



Figura 5. Página retornada pelo link.

Fonte: Dados da pesquisa – Twitter.

Viabilizado o acesso às postagens históricas, utilizaram-se técnicas de *web scraping* (Zhao, 2017), para viabilizar a coleta do texto das postagens. Apesar da complexidade aparente dos comandos utilizados, em cada página apresentada nas consultas os desenvolvedores do Twitter optaram por utilizar *tags* com semântica facilmente interpretada pela maioria dos leitores da língua inglesa (W3schools, 2021).

Vale ressaltar que essa análise descrita se refere à versão das páginas do Twitter disponíveis em janeiro/2021, podendo a empresa responsável pela mídia social alterar a forma de construção do site a qualquer momento. Na versão do site analisada, cada um dos *tweets* é inicialmente delimitado pelas *tags* HTML `<article>` e `</article>`, descritas em detalhes em https://www.w3schools.com/tags/tag_article.asp.

A linguagem de programação Python oferece algumas bibliotecas que facilitam o acesso a essas *tags*. Neste trabalho, optou-se por utilizar a biblioteca BeautifulSoup.

Para ter acesso a todas as *tags* `<article>` e seu conteúdo, basta utilizar o método *find_all* ("article"), que todas as postagens são separadas. Após essa coleta inicial, de maneira resumida, basta ignorar as *tags* CSS e JavaScript da página, para que o texto da postagem possa ser coletado. Um exemplo dessa relação entre imagem e código é apresentado nas Figuras 6 e 7.

As postagens nas mídias sociais, entretanto, não se limitam a símbolos alfanuméricos, como as letras de A a Z e os algarismos de 0 a 9. Um conteúdo semântico muito valioso a qualquer análise pode ser inferido analisando-se *emoticons*, que são pequenas imagens acrescentadas ao texto das postagens.

Neste trabalho, houve um tratamento à parte para a extração dos *emoticons*. A mídia social Twitter optou por utilizar imagens personalizadas para apresentar seus *emoticons*, muitas delas fora do padrão de comandos Unicode (Unicode®, 2021). Essas imagens são mostradas por meio da *tag* HTML ``, utilizada para carregar, para o navegador web, as imagens do servidor.

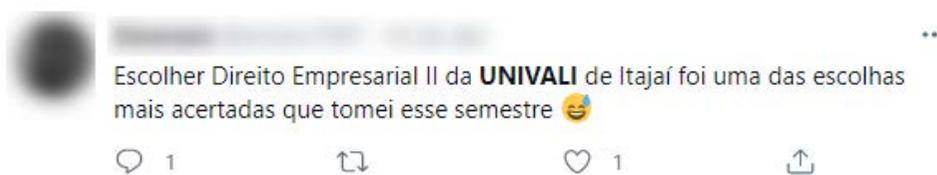


Figura 6. Tweet, conforme apresentado ao usuário pelo navegador.

Fonte: Dados da pesquisa – Twitter.

```

<article role="article" tabindex="0" class="css-1dbjc4n">

  <div lang="pt" dir="auto" class="css-9010ao">

    <span class="css-9010ao">Escolher Direito Empresarial II da </span>

      <span class="css-9010ao css-16my406">UNIVALI</span>

      <span class="css-9010ao">

        de Itajaí foi uma das escolhas mais acertadas que tomei esse semestre

      </span>

      <span>

        </span>

      </div>

    </article>

```

Figura 7. Tag <article> do tweet da Figura 5 (resumido).

Segmento do código que constrói a imagem da Figura 6. Fonte: Elaborado pelos autores.

Mesmo com essa particularidade, o Twitter informa, no atributo de texto alternativo à imagem, o caractere Unicode (Unicode®, 2021) ou a descrição da imagem. Essa extração foi realizada a partir desse atributo.

Na Figura 8, é possível visualizar o texto alternativo aos *emoticons* apresentados na Figura 5, aqui destacado em bordas escuras. No segmento sinalizado, é possível observar a descrição do *emoticon* apresentado: ‘Cara sorridente de boca aberta e suando frio’.

```

<img alt="Cara sorridente de boca aberta e suando frio" draggable="false"

  src=https://abs-0.twimg.com/emoji/v2/svg/1f605.svg class="css-9pa8cd" />

```

Figura 8. Texto alternativo ao *emoticon*, tag de imagem.

Segmento do código-fonte da página do Twitter. Fonte: Elaborado pelos autores, com base em dados da pesquisa.

Segmentação do processamento para a extração dos posts

Robôs de extração de dados podem executar uma quantidade de requisições ao site muito maior que um humano navegando normalmente no mesmo site. É comum que, por segurança, os sites deixem de responder a um volume

grande de requisições de um mesmo computador, para evitar uma sobrecarga aos servidores. No desenvolvimento do script para extração de dados, optou-se por fazer acessos com intervalos de cerca de 20 segundos entre uma requisição e outra. Conforme demonstrado na Tabela 3, estimou-se em cerca de 61 dias o tempo necessário para que todas as requisições fossem realizadas.

Tabela 3. Cálculo das requisições.

Links	Segundos entre requisições	Tempo em segundos	Tempo em minutos	Tempo em horas	Dias para requisição dos links
265.811	20	5.316.220	88603,6667	1476,7278	61,53

Nota. A quantidade de links se refere aos termos de busca utilizados. Fonte: Os autores, com base em dados da pesquisa.

Dada a estimativa inicial de cerca de 61 dias para requisição dos links, adicionados ainda ao tempo de processamento da página, ao tempo para armazenamento dos dados e ao tempo de inoperabilidade, seja por queda da internet, seja por travamento de computador ou erro e tempo de correção do script, optou-se pela segmentação da análise dos links em execução simultânea, utilizando 20 computadores.

Conforme esquema apresentado na Figura 9, os computadores foram utilizados de forma simultânea, dividindo-se entre eles o trabalho de extração dos dados. Orquestrou-se este trabalho manualmente, via computador remoto. A partir deste esquema, foi possível aumentar a capacidade de processamento na extração de dados.

Utilizando o esquema ilustrado na Figura 6, foi possível extrair cerca de 1,8 milhão de *tweets* entre os dias 10 e 20/12/2020. Todos os dados foram mantidos em arquivos e estão disponíveis para leitura e análise a serem realizadas posteriormente.

A partir da base de dados, é possível realizar diversos procedimentos técnicos, conforme o objetivo da análise. Além disso, este banco de dados pode ser lido facilmente por outros softwares, para apresentação de informações gerenciais, como análise de sentimento, polarização dos *tweets*, análises descritivas, abordagem de mineração de dados com análise exploratória de grafos, dentre outras técnicas quantitativas ou qualitativas. Na Tabela 4 são apresentados os campos extraídos para cada *tweet*.

Por meio dos termos de busca, foi possível realizar as consultas que atendessem à extração de dados desejada, a partir das quais extraiu-se o URL do *tweet*, data, nome, texto do *tweet*, *hashtags*, respostas, *retweets*, curtidas e *emoticons*. Com este processo de extração de dados é possível que pesquisadores avancem no campo das pesquisas organizacionais, devido à influência no desempenho, na resposta corporativa ao ativismo social, no impacto na reputação e na imagem institucional impulsionada pelas mídias sociais. O script desenvolvido é apresentado nos dados compartilhados deste artigo.

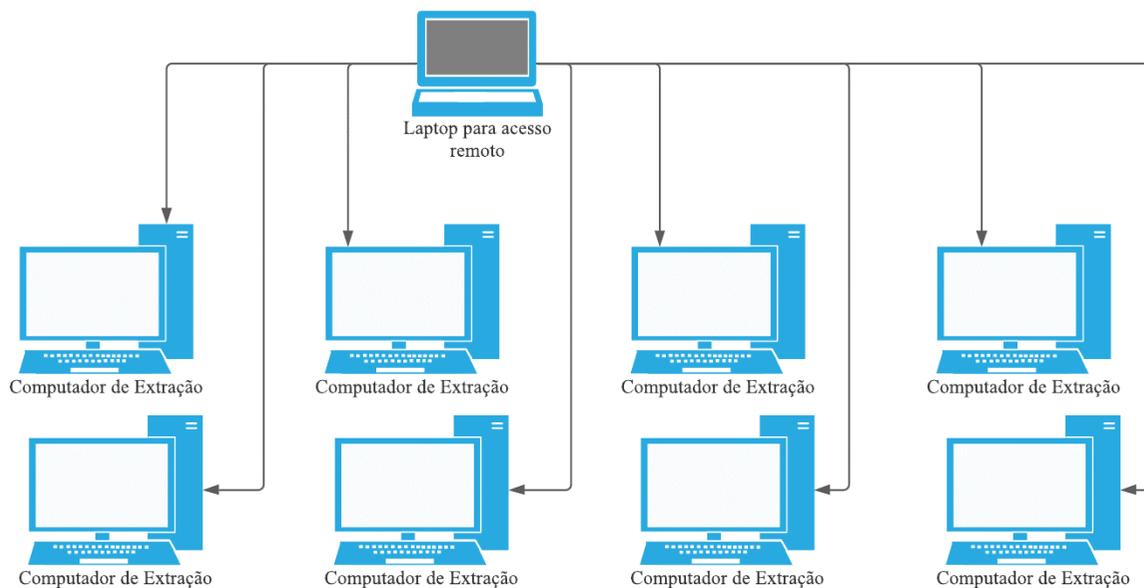


Figura 9. Esquema da segmentação do processamento.
Fonte: Elaborado pelos autores.

Tabela 4. Exemplo de dados extraídos por *tweet*.

Campos	Dados extraídos
Termo de busca	univali
url_dia	https://twitter.com/search?q=univali%20since%3A2019-12-18%20until%3A2019-12-19&src=typd
url_tweet	/hanna_coisinha/status/1207147824615493632
data	2019-12-18T03:57:23.000Z
data_descr	18 de dez de 2019
nome	@hanna_coisinha
tweet_text	O que eu mais vou sentir falta da Univali???? Os professores 42
hashtags	
resposta	4
retweet	0
curtida	2
emoticon	Cara chorando muito Coração vermelho

Nota. O *uniform resource locator* (URL) é um termo técnico que foi traduzido para a língua portuguesa como 'localizador uniforme de recursos'. Um URL se refere ao endereço de rede no qual se encontra algum recurso informático. Fonte: Os autores, com base em dados da pesquisa.

Resposta corporativa para o ativismo on-line

A partir do script desenvolvido para a extração de dados, é possível que organizações monitorarem o ambiente on-line de forma sistematizada, tendo como foco responder às demandas que surgem de forma instantânea e que possam trazer impactos à gestão organizacional, seja pelas novas estratégias de ativismo social on-line, levando ao empoderamento grupos de ativistas ao ampliar o alcance das reivindicações assumidas como direitos coletivos, seja em questões isoladas, na defesa de causas humanitárias, políticas, culturais e econômicas, mas que ao mesmo tempo exigem atenção de organizações devido ao seu potencial impacto.

Neste estudo, a partir da coleta e análise dos *tweets* de instituições de ensino superior (IESs) no Brasil foi possível evidenciar as principais categorias de influência de ativistas e *stakeholders* no ambiente on-line e, a partir disso, inferir ações estratégicas de resposta às demandas identificadas. Por fim, propõe-se a implementação do processo de monitoramento de mídia social para empresas.

Exemplo prático (estudo de caso)

Para demonstração prática a partir da seção 'Script (conjunto de comandos) para extração de dados', foram extraídos os *tweets* de uma IES do estado de Santa Catarina. A escolha da IES se deu pela sua relevância nos rankings nacionais e internacionais. A extração de dados resultou no total de 24.787 *tweets*, e a amostra aleatória extraída pelo software de 5.984 *tweets* da IES, no período de 2013 a 2019. Cabe destacar que a escolha do período teve como objetivo buscar a maior quantidade de interações com a IES em diferentes momentos. Para a análise dos dados foi utilizada a ferramenta *Orange Data Mining*. Primeiramente, foram identificadas as categorias por meio da análise de *cluster*.

A partir da frequência dos termos e pela análise de *cluster*, foram identificadas três categorias de interação com a IES: financeira, pedagógica e social. Ou seja, quando um usuário realiza um comentário sobre a IES no Twitter, seja positivo, seja negativo, este se refere a questões financeiras, pedagógicas ou sociais. Na Tabela 5 são apresentadas as palavras-chave com maior frequência, por categoria.

Tabela 5. Palavras-chave por categoria.

Palavras-chave	Categorias	N.º <i>tweets</i>	%
Mensalidade; pagar; dinheiro; multa; créditos; boleto; reais; Fies; bolsa; cobrar; gastar.	Financeira	56	0,9%
Aula; curso; estudar; palestra; prova; notas; professores; pesquisa; vestibular; matrícula; educação; formatura.	Pedagógica	1.154	19,3%
Campus; universidade; faculdade; galera; pessoas; gente; partiu; amanhã; cedo; cheguei; passei; acordar; dormir.	Social	4.033	67,4%
<i>Out of context</i>		741	12,4%
TOTAL GERAL		5.984	100%

Nota. Fonte: Elaborado pelos autores, com base em dados da pesquisa, gerados pela ferramenta *Orange Data Mining*.

Caberia, ainda, avaliar os *tweets* com maior engajamento (quantidade de curtidas e *retweets*), com vistas a identificar movimentos ativistas, por motivos individuais ou coletivos. Além disso, por meio da escala de sentimentos, é possível realizar uma análise qualitativa evidenciando os principais motivadores para questões positivas ou negativas das interações com a organização.

Implementação do processo de monitoramento de mídia social

O processo de monitoramento e gestão de mídias sociais necessita ser implementado nas empresas não apenas com a finalidade de avaliação da reputação institucional, ou para fins de comunicação e marketing, mas de uma forma mais ampla, abrangendo uma visão intersetorial. É neste ponto que o script (conjunto de comandos) para extração de dados desenvolvido neste estudo pode auxiliar na extração, categorização e distribuição das postagens, com a classificação de *tweets* positivos, negativos e neutros, conforme apresentado no exemplo prático da subseção 3.4.1.

De forma geral, o processo de monitoramento de mídias sociais é o acompanhamento contínuo e estruturado por métricas de todas as interações, reações (como curtidas, conteúdo e compartilhamentos) e comentários dos públicos, mostrando o que vem sendo dito sobre a empresa na mídia social. A visão intersetorial proposta neste artigo requer a análise ampla de todas as interações ocorridas na mídia social. Estas interações podem ser capturadas a partir dos seguintes pontos:

- o nome da marca e suas variações mais comuns, se houver;
- os nomes dos principais produtos e serviços da empresa;
- as menções, com *tags* e *hashtags* da mídia social; essas menções podem ter vários termos de busca, como o nome da marca, de produto, de serviço, de slogan de campanha lançada, entre outros;
- palavras-chave que tenham proximidade com o assunto da marca, produto ou serviço;
- menções com filtro por localização, quando for o caso;
- adicionalmente, busca por menções do principal concorrente da marca, para entender qual a percepção que o público tem de empresas do mesmo segmento.

Pelo processo de monitoramento de mídia social, a empresa poderá compreender quem são os apoiadores e os detratores, ativistas sociais, as postagens com maior engajamento, as palavras que mais são associadas à empresa e a reação do público em geral, ao longo do tempo.

A Figura 11 apresenta o fluxograma do processo de monitoramento. Este processo inicia-se com as postagens na mídia social. O robô – script (conjunto de comandos) para extração de dados desenvolvido neste estudo – realiza a captura e armazenamento das postagens em base de dados estruturada, contendo alguns campos importantes como: termo de busca; *url_dia*; *url_tweet*; *data*; *data_descr*; *nome*; *tweet_text*; *hashtags*; *resposta*; *retweet*; *curtida*; *emoticon*.

A partir do banco de dados, é possível realizar a qualificação das postagens com uma escala de alto, médio e baixo impacto. A métrica para medir o impacto pode ser baseada nas curtidas e *retweets*, ou seja, quando ocorrer uma postagem com um número significativo de curtidas e/ou *retweets*, esta é classificada como sendo de alto impacto, seja do ponto de vista positivo, seja do negativo ou neutro.

Além da classificação de impacto, a categorização dos *tweets* se faz necessária para que ocorra a distribuição das mensagens para os respectivos setores da empresa. No exemplo prático apresentado neste artigo (em uma instituição de ensino superior), a classificação se deu por três principais categorias: financeira, pedagógica e social. Todavia, estas categorias podem variar conforme o tipo da empresa, segmento, clientes, dentre outras. Desta forma, apresentam-se, no fluxograma, três categorias genéricas essenciais: preço/financeiro, qualidade e atendimento.

A categoria de preço/financeiro está associada a todas as menções que dizem respeito ao preço de produtos e/ou serviços, questões associadas a aspectos financeiros da empresa. A categoria de qualidade está relacionada ao nível em que o produto ou serviço atende aos requisitos previamente estabelecidos. Por fim, a categoria de atendimento diz respeito ao nível de satisfação de clientes e outros *stakeholders* da empresa. Apesar de esta categoria ser considerada como subjetiva, pois cada pessoa tem uma percepção distinta de qualidade, este é um pilar estratégico na captação de recursos para empresas.

Após a classificação e categorização dos *tweets*, a próxima etapa é a distribuição das informações para a análise e resposta dos setores responsáveis. A partir do estudo de caso desenvolvido nesta pesquisa, propõe-se a distribuição para cinco áreas fundamentais: (1) comercial, serviço de atendimento ao cliente – SAC e relações públicas – RP, (2) ouvidoria, (3) serviços/produção, (4) áreas administrativas e (5) gestão estratégica, podendo variar conforme a estrutura organizacional de cada empresa ou IES. Para a implementação do processo, destaca-se a importância de estruturação da área de monitoramento de mídia social. Ainda que esta atividade possa estar atrelada a setores como comercial, SAC, RP, é fundamental que a empresa direcione recursos para essa finalidade.

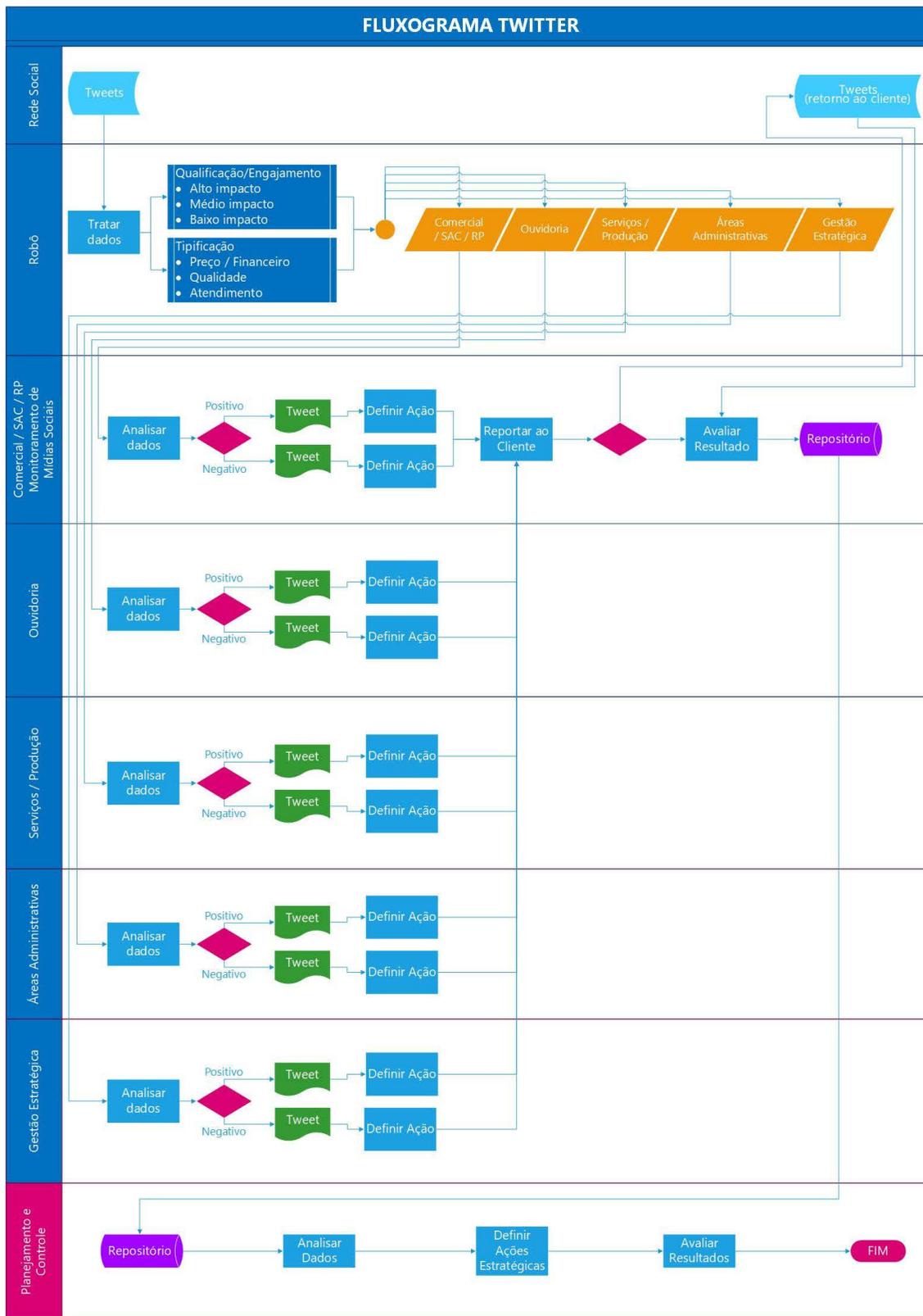


Figura 11. Fluxograma do processo de monitoramento de mídia social.

Fonte: Elaborado pelos autores, com base em dados da pesquisa.

Cada setor analisa os dados e desenvolve ações específicas, de acordo com a demanda apresentada por meio da classificação e categorização dos *tweets*. Se positivo, o setor avalia os pontos citados e remete a resposta para a mídia social. Se negativo, o setor desenvolve ação de correção ou melhoria, conforme o feedback publicado, e comunica-a aos setores comercial, SAC, RP e monitoramento de mídia social, que remete a resposta para a mídia social.

Os setores comercial, SAC, RP e monitoramento de mídia social avaliam os resultados, que são armazenados em um repositório de dados. A partir deste repositório e com as métricas de classificação e categorização dos *tweets*, a área estratégica pode analisar os dados e definir ações estratégicas. Esta última etapa do processo é fundamental para a identificação de recursos e competências necessárias e para o posicionamento estratégico da empresa.

Ações organizacionais estratégicas

O processo de extração, monitoramento e gerenciamento de mídias sociais requer um conjunto de

ações organizacionais propositivas que, se implantadas, podem contribuir com a gestão e resposta corporativa às demandas do ambiente on-line. A partir da extração de dados da web, por meio do script (conjunto de comandos) apresentado na seção 3.2, e do estudo de caso realizado, foi possível inferir um conjunto de ações de melhoria para o processo de extração de dados, monitoramento e gerenciamento das interações de usuários do Twitter com a organização. Na Tabela 7 são apresentadas estas ações.

Com a implementação das ações descritas neste estudo, é possível aprimorar o processo de monitoramento e gestão de mídias sociais tendo como premissas a melhora do engajamento de *stakeholders*, a identificação de movimentos ativistas e seguidores insatisfeitos, a detecção de informações falsas, o acompanhamento de ações da concorrência e a identificação das forças e fraquezas da organização. Para tanto, é preciso implementar o processo de monitoramento de mídia social nas empresas. Este processo requer a integração entre áreas/setores, para que ocorra a resposta corporativa de demandas advindas do ambiente on-line.

Tabela 7. Ações estratégicas de resposta ao ativismo on-line.

Etapas	Ações
Extração de dados	Automatizar o processo de extração de dados do Twitter a partir do script (conjunto de comandos) apresentado na seção 3.2 deste artigo.
	Decodificar milhares de dados (textos e imagens, incluindo <i>emojis</i>) em tempo real, de forma organizada em base de dados, para análise futura.
Monitoramento	Captar menções de usuários relacionadas à organização, com <i>tags</i> e <i>hashtags</i> , no Twitter – as <i>hashtags</i> são utilizadas como mecanismos de buscas, ferramentas de categorização e, principalmente, táticas de marketing e comunicação.
	Monitorar com dados históricos organizados de forma cronológica e acompanhamento de temas estratégicos, por meio de palavras-chave.
	Definir métricas para as interações, reações (como curtidas, conteúdo publicado e compartilhamentos) e comentários do público em geral.
Gerenciamento	Criar categorias de análise para identificação dos principais motivadores das interações de <i>stakeholders</i> com a organização (conforme demonstrado na subseção 3.4.1 – Exemplo prático).
	Identificar sentimentos positivos e negativos, ao longo do tempo, e promover ajustes organizacionais, se necessário.
	Identificar <i>tweets</i> com maior engajamento (quantidade de curtidas e <i>retweets</i>) e a quais temas estão relacionados.
	Fomentar publicações que geram impacto positivo na imagem da organização com vistas a melhorar o relacionamento organizacional com seus <i>stakeholders</i> .
	A partir de feedbacks negativos sobre determinada estratégia ou ação realizada pela organização, elaborar plano de ação para correção dos aspectos evidenciados.
	Analisar interações de <i>stakeholders</i> com organizações do setor, concorrentes, agências, ONGs e outras instituições de interesse para a organização.
	Desenvolver <i>reports</i> inteligentes a partir do monitoramento das mídias sociais, via indicadores qualitativos e quantitativos, para a gestão estratégica da organização.

Nota. Fonte: elaborado pelos autores.

CONCLUSÕES E CONTRIBUIÇÃO TECNOLÓGICA/SOCIAL

O objetivo central deste estudo foi desenvolver e estruturar uma ferramenta de gestão em mídia social. No campo prático, este estudo contribui para que as empresas

possam realizar, de forma sistematizada, a extração de dados do Twitter para o desenvolvimento de um sistema e processo robusto de monitoramento, possibilitando assim a promoção de ações estratégicas no ambiente on-line. Tal prática se justifica tendo em vista que as mídias sociais se tornaram onipresentes e estão desempenhando um papel cada vez

mais crítico nos ambientes de negócios atuais. As novas possibilidades informacionais que elas apresentam afetam as relações entre *stakeholders* e gestores de organizações pelos mais diversos motivos e interesses individuais ou coletivos.

A partir do estudo de caso apresentado, foi possível inferir um conjunto de ações para o processo de extração de dados, monitoramento e gerenciamento das interações de usuários do Twitter com a IES em estudo. A identificação das principais categorias de interação com a IES pôde auxiliar no processo de resposta às demandas do ambiente on-line.

A implementação do processo de monitoramento de mídia social requer a integração intersetores. Embora os estudos em mídias sociais concentrem atenção na área mercadológica, este estudo amplia o campo para a discussão sobre como técnicas de computação podem servir para apoio à administração e à tomada de decisão.

A partir da sistematização dos dados, é possível realizar outros estudos, que possam contribuir, por exemplo, para a literatura que examina os comportamentos estratégicos das empresas sobre a utilização do Twitter para a divulgação de informações (Davis & Thompson, 1994; Etter & Albu, 2021; Ghardallou, 2021; Rao et al., 2000; Van Dyke et al., 2004).

A evidência de que a aplicação do script (conjunto de comandos) é eficaz na extração de dados pode estimular outros pesquisadores a desenvolverem estudos na área. O monitoramento de mídias sociais permite a realização de

análise quantitativa e qualitativa, além de aplicabilidade de outros métodos que podem ajudar pesquisadores nas investigações sobre comportamentos que antes não eram identificados nos dados de mercado ou de conhecimento prévio.

No tocante à operacionalização do script para extração dos dados, notaram-se melhorias que, se realizadas, podem automatizar ainda mais o trabalho, com significativa redução da intervenção humana e com mais segurança e confiabilidade no processo de extração. Desta forma, como sugestão para estudos futuros e aprimoramentos do script apresentado neste artigo, sugere-se a utilização de um banco de dados centralizado, para armazenar os links e os *posts* extraídos de diversas máquinas, com processamento simultâneo.

O tempo de intervalo entre o acesso a um link e outro foi definido arbitrariamente, a partir de experimentações realizadas sem metodologia e catalogação. Havendo a construção do *cluster*, é possível reduzir gradativamente o tempo de espera entre um link e outro, até que seja definido um tempo mínimo de intervalo entre os links, melhorando, assim, o processo de extração dos dados. É possível, também, virtualizar os computadores em ambiente de nuvem.

Por fim, este estudo apresenta limitações na proposição das ações estratégicas para as demandas do ambiente on-line, visto que foram analisados os *tweets* de uma IES somente, o que impossibilita a generalização dos resultados.

REFERÊNCIAS

- Abdulaziz, M., Alotaibi, A., Alsolamy, M., & Alabbas, A. (2021). Topic based sentiment analysis for COVID-19 tweets. *International Journal of Advanced Computer Science and Applications*, 12(1), 626-636. <https://doi.org/10.14569/IJACSA.2021.0120172>
- Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, Auditing & Accountability Journal*, 30(4), 762-776. <https://doi.org/10.1108/AAAJ-03-2017-2880>
- Bimber, B., Flanagan, A., & Stohl, C. (2012). *Collective action in organizations: Interaction and engagement in an era of technological change*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1080/10584609.2012.722513>
- Briscoe, F., & Gupta, A. (2016). Social activism in and around organizations. *Academy of Management Annals*, 10(1), 671-727. <https://doi.org/10.1080/19416520.2016.1153261>
- Cavalcanti, D. B., Bringel, E. P. B., Costa, F. R. J. da, Oliveira, T. M. de, & Zuccolotto, V. R. (2019). Digital activism and indignation nets in Brazil: The pressure groups. *Journal of Politics in Latin America*, 11(1), 109-130. <https://doi.org/10.1177/1866802X19840455>
- Choudhary, P. (2021). *TwitterAdvSearch*. Retrieved from <https://github.com/prakharchoudhary/TwitterAdvSearch>
- Davis, G. F., & Thompson, T. A. (1994). A social movement perspective on corporate control. *Administrative Science Quarterly*, 39(1), 141-173. <https://doi.org/10.2307/2393497>
- D'Andréa, C. (2021). Para além dos dados coletados: Políticas das APIs nas plataformas de mídias digitais. *Matrizes*, 15(1), 103-122. <https://doi.org/10.11606/issn.1982-8160.v15i1p103-122>
- D'Haen, J., Van den Poel, D., Thorleuchter, D., & Benoit, D. (2016). Integrating expert knowledge and multilingual web crawling data in a lead qualification system. *Decision Support Systems*, 82, 69-78. <https://doi.org/10.1016/j.dss.2015.12.002>
- Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the internet age*. Cambridge, MA: MIT Press.
- Etter, M., & Albu, O. B. (2021). Activists in the dark: Social media algorithms and collective action in two social movement organizations. *Organization*, 28(1), 68-91. <https://doi.org/10.1177/1350508420961532>

- Fundação Instituto de Administração (2021). *Ativismo nas redes sociais: características, impactos e exemplos*. Retrieved from <https://fia.com.br/blog/ativismo-nas-redes-sociais>
- George, J. J., & Leidner, D. E. (2019). From clicktivism to hacktivism: Understanding digital activism. *Information and Organization*, 29(3), 100249. <https://doi.org/10.1016/j.infoandorg.2019.04.001>
- Ghardallou, W. (2021). The impact of firms' and CEOs' social media usage on corporate performance. *Investment Management & Financial Innovations*, 18(4), 21. [https://doi.org/10.21511/imfi.18\(4\).2021.03](https://doi.org/10.21511/imfi.18(4).2021.03)
- Gomez-Carrasco, P., & Michelon, G. (2017). The power of stakeholders' voice: The effects of social media activism on stock markets. *Business Strategy and the Environment*, 26(6), 855–872. <https://doi.org/10.1002/bse.1973>
- Harquail, C. (2011). Recreating reputation through authentic interaction: Using social media to connect with individual stakeholders. In R. Burke, G. Martin, & C. Cooper, (Eds.), *Corporate reputation. Managing opportunities and threats* (pp. 245-263). Burlington, NJ: Gower Publishing.
- Harris, A. L., & Rea, A. (2009). Web 2.0 and virtual world technologies: A growing impact on IS education. *Journal of Information Systems Education*, 20(2), 137. Retrieved from <https://aisel.aisnet.org/jisc/vol20/iss2/3>
- Instituto Brasileiro de Pesquisas e Análise de Dados (2021). *Twitter libera API especial para acadêmicos: O que você precisa saber*. Retrieved from <https://www.ibpad.com.br/blog/twitter-libera-api-especial-para-academicos-o-que-voce-precisa-saber/>
- Jenzen, O., Erhart, I., Eslen-Ziya, H., Korkut, U., & McGarry, A. (2021). The symbol of social media in contemporary protest: Twitter and the Gezi Park movement. *Convergence*, 27(2), 414-437. <https://doi.org/10.1177/1354856520933747>
- Jung, M. J., Naughton, J. P., Tahoun, A., & Wang, C. (2018). Do firms strategically disseminate? Evidence from corporate use of social media. *The Accounting Review*, 93(4), 225-252 <https://doi.org/10.2308/accr-51906>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Khalil, S., & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, 6, 98-106. <https://doi.org/10.1016/j.softx.2017.04.004>
- Kim, E. H., & Youm, Y. N. (2017). How do social media affect analyst stock recommendations? Evidence from S&P 500 electric power companies' Twitter accounts. *Strategic Management Journal*, 38(13), 2599-2622. <https://doi.org/10.1002/smj.2678>
- Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, 53(2), 367–404. <https://doi.org/10.1111/1475-679X.12074>
- Lei N° 13.709 de 14 de agosto de 2018 (2018). Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União de 15 de agosto de 2018*, e republicado parcialmente em 15 de agosto de 2018 - Edição extra. Retrieved from www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/113709.htm
- Lewis, C. (2005). Unions and cyber-activism in South Africa. *Critical Perspectives on International Business*, 1(2/3), 194–208. <https://doi.org/10.1108/17422040510595672>
- Luo, X. R., Zhang, J., & Marquis, C. (2016). Mobilization in the internet age: Internet activism and corporate response. *Academy of Management Journal*, 59(6), 2045-2068. <https://doi.org/10.5465/amj.2015.0693>
- Meirelles, P. (2019). *Histórico das APIs no monitoramento e pesquisa em mídias sociais*. Brasília, DF: Instituto Brasileiro de Pesquisa e Análise de Dados.
- Mercado, L. P. L., & da Silva, I. P. (2013). O Twitter na coleta de dados na pesquisa qualitativa. *Informática na educação: Teoria & Prática*, 16(1). <https://doi.org/10.22456/1982-1654.12378>
- Omari, A., Shoham, S., & Yahav, E. (2016). Cross-supervised synthesis of web-crawlers. In 38th *Proceedings of the International Conference on Software Engineering*, (pp. 368-379). New York, NY: ACM. <https://doi.org/10.1145/2884781.2884842>
- Rao, H., Morrill, C., & Zald, M. N. (2000). Power plays: How social movements and collective action create new organizational forms. *Research in Organizational Behavior*, 22, 237–281. [https://doi.org/10.1016/S0191-3085\(00\)22007-8](https://doi.org/10.1016/S0191-3085(00)22007-8)
- Ross, S. D., & Cruz, B. D. P. A. (2021). Análise quantitativa de textos: Apresentação e operacionalização da técnica via Twitter. *Administração: Ensino e Pesquisa*, 22(1). <https://doi.org/10.13058/raep.2021.v22n1.1859>
- Selenium (2021). *Selenium Automates Browsers*. Retrieved from <https://www.selenium.dev/>
- Shin, S., & Ki, E.-J. (2022). Understanding environmental tweets of for-profits and nonprofits and their effects on user responses. *Management Decision*, 60(1), 211-230. <https://doi.org/10.1108/MD-05-2020-0679>
- Silva, T., & Stabile, M. (2016). *Análise de redes em Mídias Sociais*. In Monitoramento e pesquisa em mídias sociais: Metodologias, aplicações e inovações (pp. 237-260). São Paulo: Uva Limão.
- Smits, M., & Mogos, S. (2013). *The impact of social media on business performance*. 21st Proceedings of the European Conference on Information Systems. p. 125. Retrieved from https://www.researchgate.net/publication/273138667_The_impact_of_social_media_on_business_performance
- Statista (2021). *Forecast of the number of Twitter users in Brazil from 2017 to 2025*. Retrieved from www.statista.com/forecasts/1146589/twitter-users-in-brazil
- Twitter (2022a). *Política de privacidade do Twitter*. Retrieved from <https://twitter.com/pt/privacy>

- Twitter (2022b). *Robots*. Retrieved from <https://twitter.com/robots.txt>
- Unicode® (2021). *Full emoji list, v14.0*. Unicode® Emoji Charts v14.0. Retrieved from <https://unicode.org/emoji/charts/full-emoji-list.html>
- Van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, 1(1), 2-14. Recuperado Retrieved from www.researchgate.net/publication/263566996_Understanding_Social_Media_Logic
- Van Dyke, N., Soule, S. A., & Taylor, V. A. (2004). The targets of social movements: Beyond a focus on the state. *Research in Social Movements, Conflicts, and Change*, 25(1), 27-51. Retrieved from [www.emerald.com/insight/content/doi/10.1016/S0163-786X\(04\)25002-9/full/html](http://www.emerald.com/insight/content/doi/10.1016/S0163-786X(04)25002-9/full/html)
- Vogler, D. (2020). Analyzing reputation of Swiss universities on Twitter: The role of stakeholders, content and sources. *Corporate Communications: An International Journal*,

25(3), 429-445. Retrieved from <https://www.emerald.com/insight/content/doi/10.1108/CCIJ-04-2019-0043/full/html>

- Wetzstein, B., Leitner, P., Rosenberg, F., Dustdar, S., & Leymann, F. (2011). Identifying influential factors of business process performance using dependency analysis. *Enterprise Information Systems*, 5(1), 79-98. <https://doi.org/10.1080/17517575.2010.493956>
- W3schools. (2021). *On-line web tutorials*. Retrieved from <https://www.w3schools.com/css/>
- Zhang, J., & Luo, X. R. (2013). Dared to care: Organizational vulnerability, institutional logics, and MNCs' social responsiveness in emerging markets. *Organization Science*, 24(6), 1742-1764. <https://doi.org/10.1287/orsc.1120.0813>
- Zhao, B. (2017). Web scraping. In L. Schintler, & C. McNeely (Eds.). *Encyclopedia of Big Data*. New York: Springer.

Autoria

Djeison Siedschlag*

Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração
Rua João Coan, n. 400, Universitários, 88161-064, Biguaçu, SC, Brasil
E-mail: djeison.s@univali.br

 <https://orcid.org/0000-0001-6986-573X>

Jeferson Lana

Universidade do Vale do Itajaí
R. Uruguai, n. 458, Centro 88302-901, Itajaí, SC, Brasil
E-mail: jlana@univali.br

 <https://orcid.org/0000-0002-9787-1114>

Roberto Gonçalves Augusto Junior

Universidade do Vale do Itajaí
R. Uruguai, n. 458, Centro 88302-901, Itajaí, SC, Brasil
E-mail: roberto.augusto@edu.univali.br

 <https://orcid.org/0000-0002-8735-931X>

Rosilene Marcon

Universidade do Vale do Itajaí, Programa de Pós-graduação em Administração
Rua João Coan, n. 400, Universitários, 88161-064, Biguaçu, SC, Brasil
E-mail: rmarcon@univali.br

 <https://orcid.org/0000-0002-0478-7715>

* Autor Correspondente

Conflito de Interesses

Os autores informaram que não há conflito de interesses.

Contribuições dos Autores

1º autor: conceituação (igual); administração de projeto (igual); recursos (liderança); supervisão (liderança); validação (liderança); visualização (igual); escrita - rascunho original (liderança).

2º autor: análise formal (igual); metodologia (igual); recursos (suporte); software (igual); validação (igual); visualização (igual); escrita - rascunho original (igual).

3º autor: conceituação (igual); supervisão (igual); validação (igual); escrita - revisão e edição (igual).

4ª autora: conceituação (igual); supervisão (igual); validação (igual); escrita - revisão e edição (igual)

Disponibilidade dos Dados

Os autores afirmam que todos os dados utilizados na pesquisa foram disponibilizados publicamente, e podem ser acessados por meio da plataforma Harvard Dataverse:



Siedschlag, Djeison; Augusto Junior, Roberto Gonçalves; Lana, Jeferson; Marcon, Rosilene, 2022, "Replication Data for: "Like, share and react: Twitter capture for research and corporate decisions" published by RAC-Revista de Administração Contemporânea", Harvard Dataverse, V1. <https://doi.org/10.7910/DVN/UZUJL3>

A RAC incentiva o compartilhamento de dados mas, por observância a ditames éticos, não demanda a divulgação de qualquer meio de identificação de sujeitos de pesquisa, preservando a privacidade dos sujeitos de pesquisa. A prática de *open data* é viabilizar a reproducibilidade de resultados, e assegurar a irrestrita transparência dos resultados da pesquisa publicada, sem que seja demandada a identidade de sujeitos de pesquisa.

Direitos Autorais

A RAC detém os direitos autorais deste conteúdo.

Financiamento

Os autores relataram que não houve suporte financeiro para a pesquisa deste artigo.

Verificação de Plágio

A RAC mantém a prática de submeter todos os documentos aprovados para publicação à verificação de plágio, mediante o emprego de ferramentas específicas, e.g.: iThenticate.

Método de Revisão por Pares

Este conteúdo foi avaliado utilizando o processo de revisão por pares duplo-cego (*double-blind peer-review*). A divulgação das informações dos pareceristas constantes na primeira página e do Relatório de Revisão por Pares (Peer Review Report) é feita somente após a conclusão do processo avaliativo, e com o consentimento voluntário dos respectivos pareceristas e autores.

A RAC é membro e subscreve os princípios do COPE - Comitê de Ética em Publicações

