



## Projeto de Graduação em Computação III

SISTEMA DE COLETA DE DADOS NA INTERNET PARA MEDIÇÃO DE REPUTAÇÃO DE  
EMPRESAS EM ARQUITETURA DISTRIBUÍDA

Orientador: Dr. Francisco Isidro Massetto

Aluno: Lucas Andrade Cioffi

Santo André – SP

Abril de 2021

## **1 - INTRODUÇÃO**

A imagem pública de uma pessoa, grupo, empresa ou organização pode ser um de seus ativos mais valiosos.

Nos dias atuais, a grande maioria das empresas presentes na bolsa de valores possuem uma presença em redes sociais. Redes sociais têm um impacto em todas as esferas de performance do negócio<sup>[1]</sup>. Empresas com boas reputações têm mais facilidade em encontrar e manter clientes e em anunciar seus produtos. O oposto é válido para empresas com reputações ruins<sup>[2]</sup>. A percepção do consumidor sobre a imagem da marca é fundamental na decisão de compra ou na escolha de uma empresa em favor de outras<sup>[3]</sup>.

A imagem pública também é imensamente importante para indivíduos famosos (também conhecidos como influencers) e organizações no geral, pois ambos dependem direta ou indiretamente de seu alcance e percepção por parte do público. Notícias, escândalos ou até mesmo informações falsas podem prejudicar tais percepções públicas e afetar seus objetivos negativamente. Por outro lado, notícias positivas ou campanhas de marketing podem fazer com que o público tenha uma imagem mais idônea do sujeito ou organização e podem expandir o alcance de suas comunicações.

Em uma pesquisa com executivos de relações públicas, um dos desafios relacionados a redes sociais que foi frequentemente citado é a falta de controle sobre o que as pessoas podem fazer ou dizer dentro das redes – isso inclui a propagação de informações falsas, criticismo e ações por parte de grupos ativistas<sup>[4]</sup>. Em outras palavras, há uma ameaça de propagação de informações danosas, que gera um risco para a imagem pública da empresa. Nenhuma empresa possui controle completo sobre a internet, suas redes sociais ou o que os usuários podem postar nelas, então mitigar completamente esse risco é impossível.

No entanto, ainda é interessante que sujeitos ou empresas tentem mitigar esse risco possuindo frentes dedicadas a relações públicas e que também monitorem os espaços de discussão públicos na internet. Essa monitoração permite que o sujeito ou empresa tenha conhecimento do que está sendo falado sobre si também fique ciente o mais cedo possível que uma informação danosa foi postada ou está sendo circulada. Dessa forma, a empresa ou o sujeito pode tomar ações tempestivas para amenizar os prejuízos e defender sua imagem pública ou reputação.

No cenário cibernético atual, os espaços de discussão onde as reputações de uma organização ou indivíduo podem ser discutidos são extremamente numerosos e volumosos.

Alguns exemplos, dentre outros, são: Redes sociais como Twitter, Facebook e Instagram; Fóruns, anonimizados ou não, como Reddit; Grupos e conversas diretas em aplicativos de comunicação como WhatsApp e Telegram; Notícias e comentários em notícias, como Uol; Vídeos e comentários em vídeos, como no Youtube; Sites de reviews e avaliações públicas, como Google Maps ou TripAdvisor.

Monitorar tais canais de discussão manualmente por menções a um indivíduo ou organização, nem que seja apenas um canal, e transformar tais menções em indicadores de reputação e outras formas de inteligência seria extremamente custoso ou praticamente impossível.

Neste trabalho, propomos uma solução open-source, gratuita e em arquitetura distribuída para monitorar e coletar menções a qualquer alvo (organização, pessoa ou tema de interesse). Estatísticas, relatórios e outros dados e conclusões interessantes são gerados a partir das menções coletadas. Acreditamos que essa solução poderia ser usada para monitorar a percepção pública e assuntos comentados sobre o alvo com facilidade.

### 1.1 – Justificativa

Atualmente já existem soluções de monitoração de reputação ou de presença digital, mas estas são majoritariamente proprietárias e com custo. Um exemplo de tal solução é o Mention, que permite monitorar menções à marca, possui filtragem e busca dentre as menções coletadas e pode exibir visualizações sobre a imagem pública. O plano mais básico dessa solução custa 29 USD por mês<sup>[5]</sup>, como mostrado na imagem a seguir.

**Figura 01** – Tabela de preços do Mention

### Choose your plan

There's a plan for every size and type of team - choose the one that fits your online monitoring and social media publishing goals. Feel free to contact us for more information.

Yearly (Save 2 months) ☐ Monthly ☒


Solo	Pro	ProPlus	Company
<b>\$29/mo</b>	<b>\$99/mo</b>	<b>\$199/mo</b>	<b>\$450+/mo</b>
For entrepreneurs and individuals	For marketers and small-businesses	For growing businesses and skilled-professionals	For teams and agencies
<ul style="list-style-type: none"><li>✓ 2 Basic alerts</li><li>✓ 5,000 mentions</li><li>✓ 5 social accounts</li><li>✓ 1 user</li></ul>	<ul style="list-style-type: none"><li>✓ 5 Basic alerts</li><li>✓ 10,000 mentions</li><li>✓ 10 social accounts</li><li>✓ Unlimited users</li></ul>	<ul style="list-style-type: none"><li>✓ 7 Standard alerts</li><li>✓ 20,000 mentions</li><li>✓ 15 social accounts</li><li>✓ Unlimited users</li></ul>	<ul style="list-style-type: none"><li>✓ 10 Advanced alerts</li><li>✓ 100,000+ mentions</li><li>✓ Unlimited social accounts</li><li>✓ Unlimited users</li></ul>
<a href="#">Start free trial</a>	<a href="#">Start free trial</a>	<a href="#">Start free trial</a>	<a href="#">Get a demo</a>

Fonte: <https://mention.com/en/pricing/>. Acesso em 18 mar. 2021

Outras soluções pagas e proprietárias de monitoração de marca são: Alerti, Marketing360 e AgoraPulse. A imagem abaixo mostra as funcionalidades anunciadas pelo AgoraPulse: observação de menções à marca, capacidade de resposta a conversas mencionando a marca e ações à menções como retweets, respostas, mensagens diretas e mais<sup>[6]</sup>.


**Figura 02** – Funcionalidades do AgoraPulse

## Stay informed and stay ahead




### Listen

in on social conversations about your brand, competition and space.



### Respond

quickly to urgent conversations mentioning your brand.



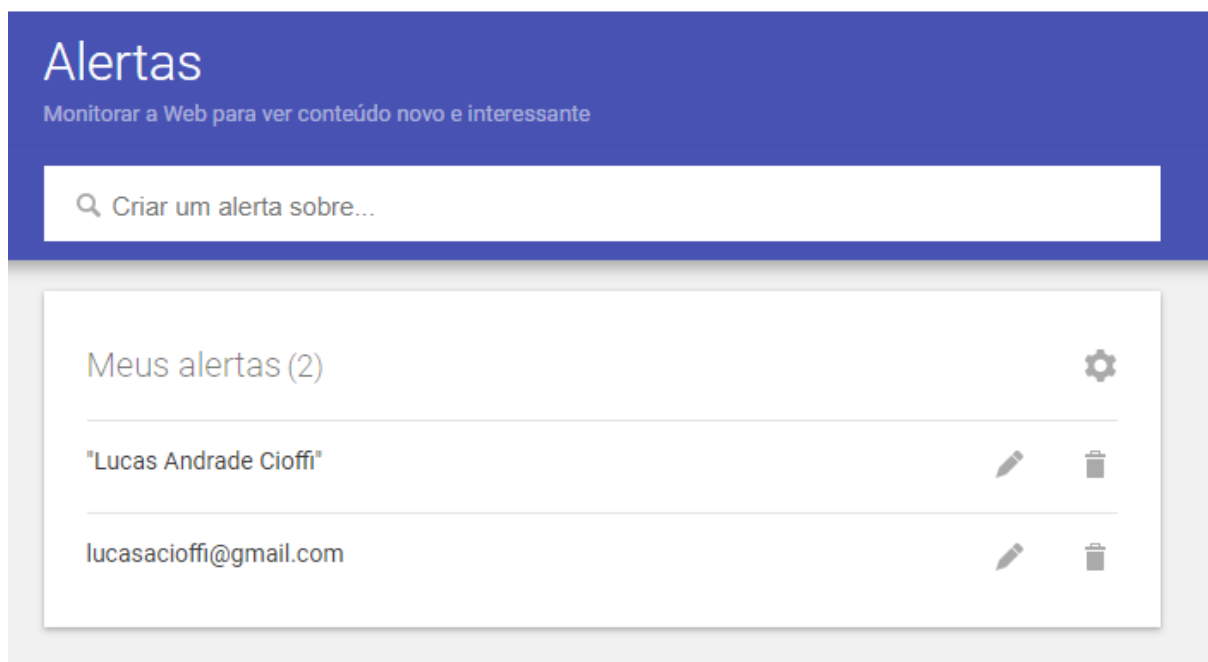
### Take action

with retweets, replies, DMs/PMs and more.

Fonte: <https://www.agorapulse.com/features/social-media-monitoring/>. Acesso em 18 mar. 2021

Existem também ferramentas gratuitas, mas elas não fornecem uma visão completa ou um histórico da reputação, apenas alertas sobre conteúdos novos publicados. Um exemplo de tais soluções é o Google Alerts, que permite que um usuário cadastre um termo qualquer e receba um alerta sempre que esse termo for mencionado em algum resultado de busca do Google<sup>[7]</sup>:

**Figura 03** – Alertas do Google



Fonte: Autor

No entanto, esses alertas não são retroativos (ou seja, não são coletadas menções anteriores à criação do alerta) e eles monitoram apenas os resultados do próprio Google. Além disso, o único formato de alerta disponível é por e-mail.

Hoje não existe uma solução para monitoração de imagem pública completamente gratuita e open-source com as mesmas funcionalidades que uma solução proprietária e com custo. Monitorar e acompanhar manualmente a imagem pública de um alvo é praticamente impossível, considerando o volume de mensagens que são postadas em qualquer rede social diariamente.

Desta forma, considerando a inviabilidade de monitorar manualmente a imagem pública de um alvo, mas a tremenda importância de tal imagem no mundo moderno, acreditamos que se torna extremamente vantajosa e útil a existência de uma ferramenta que possa realizar essa monitoração de forma automatizada e sem custo.

Como essa ferramenta deve servir o bem público, ela deve ser de fácil instalação e utilização, gratuita, open-source e modular para que outros contribuidores possam expandi-la ou modificá-la de acordo com suas necessidades. Ela deve utilizar conceitos de arquitetura distribuída para tornar-se escalável; desta forma, em casos em que há muito volume de menções ou onde o usuário queira buscar menções a diversos termos e sujeitos diferentes, ele poderá facilmente aumentar a capacidade da solução.

## 1.2– Objetivos

O objetivo técnico deste trabalho é desenvolver uma solução de monitoração composta por uma extensão open-source para a versão gratuita do Splunk<sup>[8]</sup> e uma série de coletores de dados em arquitetura distribuída que, juntos, devem monitorar a imagem pública de um sujeito, organização, instituição ou qualquer outro objeto e apresentar ao usuário final uma gama de informações úteis. Falaremos mais sobre o Splunk no próximo capítulo, mas por ora vale entender que ele é um banco de dados não relacional extensível.

Os coletores de dados devem buscar por menções a termos ou sujeitos em fontes diferentes e enviar os dados brutos para a extensão do Splunk. Eles devem utilizar conceitos de arquitetura distribuída para serem facilmente escaláveis - isso é - se o usuário desejar monitorar termos ou fontes adicionais, ele poderá facilmente adicionar mais coletores. Para este trabalho, criaremos apenas coletores da rede social Twitter.

A extensão do Splunk, por sua vez, é responsável por receber os dados, armazená-los, processá-los e transformá-los em inteligência útil para o usuário. Ela deve produzir relatórios e ter um painel de usuário que apresente ao usuário os indicadores sobre os dados coletados, informações importantes sobre a imagem pública do sujeito monitorado e quaisquer insights gerados.

O objetivo final é fornecer aos usuários uma visão automaticamente mantida da imagem pública dos alvos de interesse: o que está sendo falado sobre os sujeitos de interesse, por quais pessoas, qual o volume de menções, qual o histórico de sentimentos sobre os sujeitos de interesse e outros dados relevantes. Tais informações podem apoiar na tomada de decisões relacionadas à marketing, negócios, comunicação, atendimento ao cliente e afins.

Tal solução deve ser open-source, gratuita, modular, escalável e de fácil utilização.

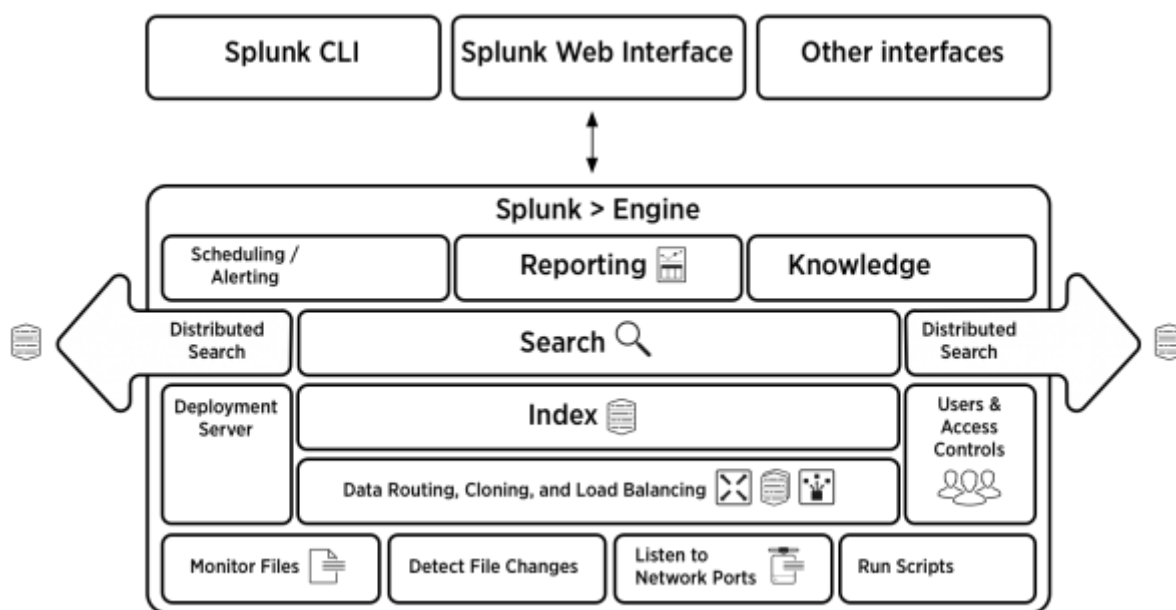
## 2 – REVISÃO BIBLIOGRÁFICA

Para a criação dessa solução pretendemos utilizar a ferramenta Splunk em conjunto à uma rede de coletores de informações (menções aos sujeitos monitorados) em arquitetura distribuída. A solução inteira pode facilmente ser hospedada na nuvem, se o usuário desejar, pois não depende da presença de um servidor físico e o Splunk possui uma versão SaaS (Software as a Service, onde o fabricante hospeda a solução para você).

O Splunk é um banco de dados não relacional proprietário, mas com uma versão gratuita que suporta até 500 Megabytes de ingestão de dados por dia. Tal volume deve ser suficiente para indivíduos, influencers ou organizações pequenas.

O diagrama abaixo mostra a arquitetura geral do Splunk:

**Figura 04** – Arquitetura geral do Splunk



Fonte: <https://docs.splunk.com/images/thumb/2/26/Architecture-new.png/600px-Architecture-new.png>. Acesso em 19 mar. 2021

Como pode-se ver, o Splunk é o motor de indexação e busca que pode ser acessado por uma interface CLI (para tarefas programáticas), por uma interface web (para usuários comuns) e por outras interfaces. Dentro dele existem diversos módulos diferentes e interconectados. Ele possui funções, por exemplo, para monitorar arquivos e registrar mudanças, receber dados via portas de rede, executar scripts Python, gerenciar usuários e permissões dentro da plataforma,

gerar relatórios e alertas com base nos dados ingeridos, e o mais importante: o recebimento, indexação e posterior busca, recuperação, filtragem e processamento desses dados.

O Splunk como um todo é escalável por design. Ele possui diversas arquiteturas de instalação com níveis diferentes de suporte de carga, redundância e velocidade. A mais simples é o aplicativo do Splunk instalado em um computador genérico, mas existem arquiteturas multi-servidor com hardware dedicado, servidores exclusivos para funções críticas, alto desempenho, hospedagem na nuvem, suporte para múltiplos usuários e redundância. O usuário fica livre para utilizar a arquitetura que julgar necessária para sustentar a carga desejada ou seus requisitos jurídicos, de governança, de auditoria ou de confiabilidade.

Todas as funções responsáveis pela ingestão de dados, por exemplo, podem ser delegadas para servidores específicos para isso. Tais tipos de mudança são úteis quando o servidor principal começa a ter problemas de sobrecarga ou o usuário deseja uma arquitetura mais redundante e escalável. O mesmo pode ser feito com diferentes módulos como o de indexação de dados ou o de busca e processamento.

O Splunk também já fornece um framework para recepção, coleta, busca e filtragem de dados. É extremamente fácil inserir dados no Splunk e depois filtrá-los e combiná-los conforme o necessário para gerar relatórios, alertas, insights e afins utilizando uma linguagem poderosa e própria da ferramenta chamada SPL (Splunk Processing Language). Falaremos mais sobre a SPL ao longo deste trabalho.

O Splunk possui diversas formas de recebimento de dados. Alguns exemplos são: envio de dados via chamadas de API; Envio de dados por requisições HTTP ou HTTPS; Envio de dados por túneis TCP ou UDP; Monitoração de arquivos na máquina como logs ou arquivos de texto, enviando as mudanças como dados ou uma cópia inteira do arquivo em uma periodicidade configurada.

O uso do Splunk facilita uma grande porção do projeto e nos permite focar melhor em nossos objetivos, mas sem comprometer as restrições apontadas na seção de objetivos.

A parte da coleta de dados será feita por meio de robôs especializados e que podem ser executados de maneira assíncrona, distribuída e elástica. O objetivo principal deles é coletar os dados desejados, realizar quaisquer filtragens ou pré-processamentos necessários e então enviá-los para o Splunk. O Splunk, por sua vez, é responsável por receber, armazenar e processar os dados captados pelos robôs.

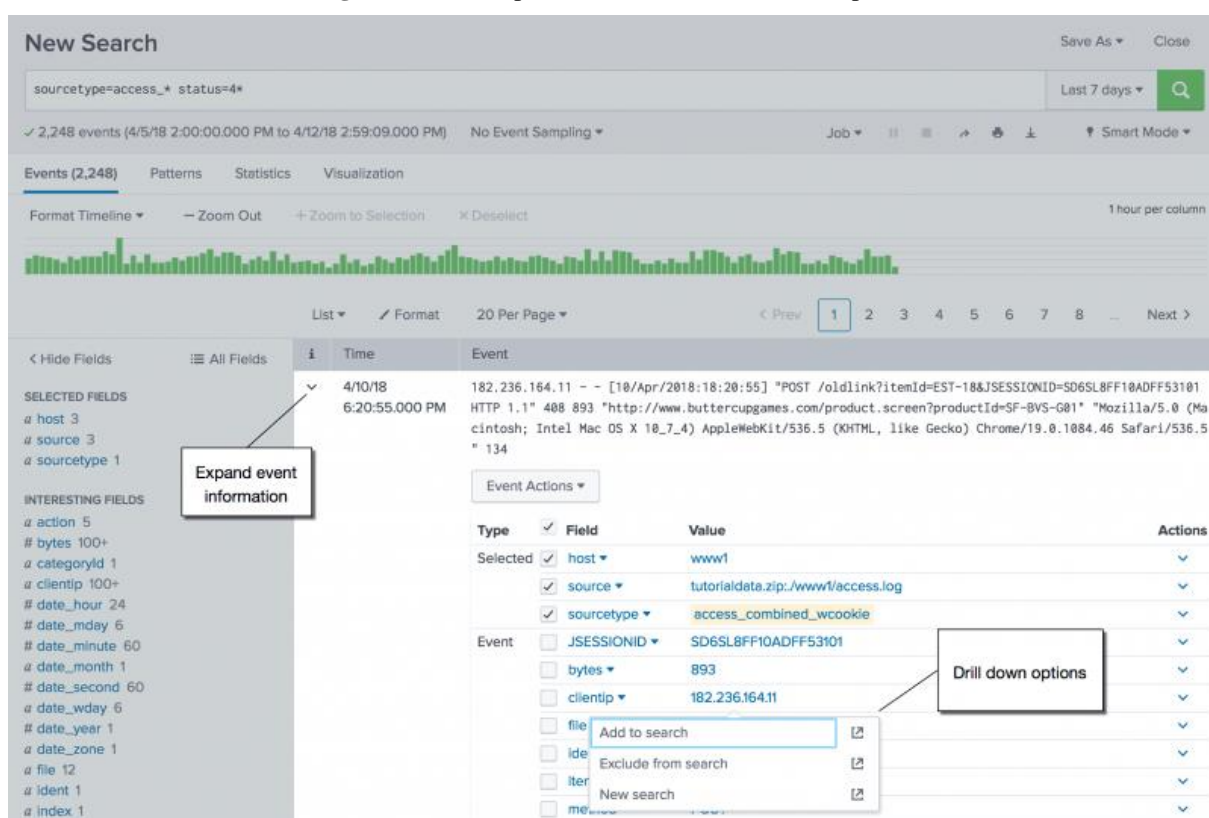
O Splunk separa os dados recebidos em “eventos” distintos. Um evento para o Splunk é apenas um bloco de texto com uma data associada. Essa data pode vir do próprio bloco de texto ou, por padrão, é utilizada a data de recebimento do texto no Splunk. Cabe ao usuário



decidir onde fica a divisão entre um evento e outro, ou qual será a data utilizada para catalogar o evento.

No momento em que o Splunk recebe os dados, ele realiza um processamento inicial para catalogar os “campos” do evento. Tais campos são, por exemplo: o conteúdo do tweet, o nome do autor, o número de citações, o número de likes e afins. Após esse processamento eles ficam disponíveis para buscas e processamentos adicionais. A imagem abaixo mostra um exemplo de como o Splunk armazena um evento e os seus metadados:

**Figura 05** – Exemplo de evento e metadados do Splunk



Fonte: <https://docs.splunk.com/Documentation/Splunk/8.1.2/Search/Drilldownnoneventdetails>. Acesso em 19 mar. 2021

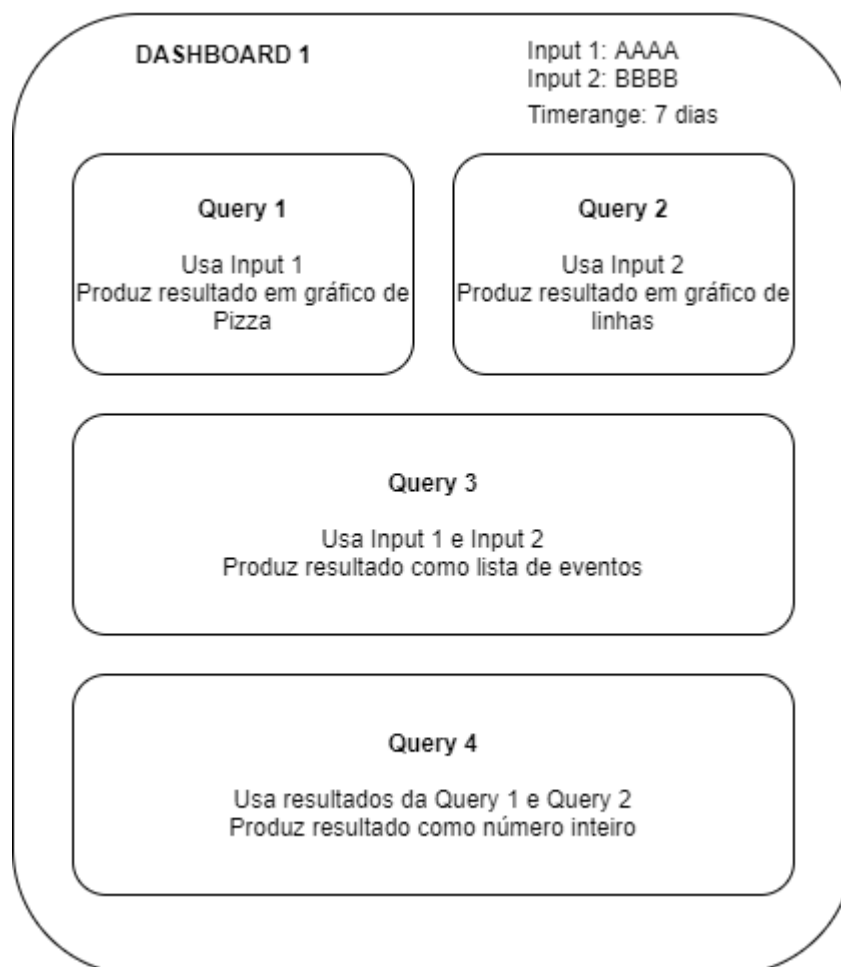
O processamento dos dados e geração de resultados é feito por meio de “queries”, que são as buscas feitas em SPL (Splunk Processing Language), uma linguagem própria do Splunk.

A linguagem SPL é semelhante à linguagem SQL, mas com funcionalidades de linguagens imperativas e a sintaxe de pipes de sistemas Unix. A linguagem SPL é capaz de realizar filtragens, buscas, tabulações, operações estatísticas e afins assim como na linguagem SQL, mas também é capaz de realizar operações de linguagens imperativas como por exemplo definir e manipular variáveis, tomar decisões e criar loops. Assim como nos pipes em sistemas

Unix, cada comando está separado por um pipe (uma barra vertical, “|”), que indica que o resultado do comando anterior será passado como entrada do comando seguinte. A saída final do programa é a saída do último comando.

Um conjunto de queries dentro do Splunk podem ser agrupadas e apresentadas em um dashboard, onde é possível também adicionar elementos como inputs para todas as queries. A figura abaixo mostra a estrutura lógica de um dashboard.

**Figura 06** – Estrutura de um dashboard do Splunk



Fonte: Autor

As imagens abaixo mostram o resultado de algumas queries criadas para a dashboard principal do projeto:

**Figura 07 – Pannel com tweets mais recentes**

Tweets mais recentes		
Tweets mais recentes		
_time ↕	tweet ↕	link ↕
2021-01-17 19:25:11	@PabloSangalli @cafecomferri @AprendaComTC Siiim!!!	https://twitter.com/FachettiEliane/status/1350932203480510471
2021-01-17 17:02:18	Está valendo! O Desafio do Milhão, com @cafecomferri já começou! E as operações com opções não param. Pela primeira vez dentro do @AprendaComTC, Ferri vai tentar transformar R\$100 mil em 1 milhão. Quer a planilha de resultados das operações 100% grátis? https://t.co/NIU0aY9ZWX	https://twitter.com/PabloSangalli/status/1350896245490003970
2021-01-17 00:40:40	@educhamps @HugoQueiroz_ @ThalisYves @AprendaComTC DEAD3 ?! 🤔🤔	https://twitter.com/AndreCappai/status/1350649210539044869
2021-01-17 00:39:52	@ThalisYves @HugoQueiroz_ @AprendaComTC DEAD3 ?! 🤔🤔	https://twitter.com/AndreCappai/status/135064908205795330
2021-01-16 21:52:37	@HugoQueiroz_ @ThalisYves @AprendaComTC @HugoQueiroz_ bem interessante esse possível IPO da Cortel. Líder de um setor pouco explorado no BR. Extremamente rentável (pelo q sei), e realmente pouco intensivo em capital, e com grande possibilidade de receita recorrente e Cross selling. Gostaria mto de ver no Matrix. Absss	https://twitter.com/educhamps/status/1350606919166263297
2021-01-16 21:25:52	@HugoQueiroz_ @cafecomferri @AprendaComTC PARabens	https://twitter.com/luizthiagos/status/1350600188713443328
2021-01-16 20:26:40	@HugoQueiroz_ @cafecomferri @AprendaComTC Show!	https://twitter.com/juliovieiradc/status/1350585292256903169
2021-01-16 20:26:29	@juliovieiradc @cafecomferri @AprendaComTC Tradeideas	https://twitter.com/HugoQueiroz_/status/1350585244295061505
2021-01-16 20:13:11	@HugoQueiroz_ @cafecomferri @AprendaComTC Hugo, quando tu fala "aprendacomtc" foi no seu trade ideas ou tem um painel específico?	https://twitter.com/juliovieiradc/status/1350581898863587334
2021-01-16 16:28:22	Agora sim se iniciou o final de semana! 🤔🤔 Obrigado pela audiência das Masters/Privatees no @AprendaComTC e dos alunos no curso de AT @tcschoolbr ! https://t.co/btNjVeJ0n3	https://twitter.com/sergiosanita/status/1350525319187456006
<div> <div>« Prev</div> <div>1</div> <div>2 3 4 Next »</div> </div> <div> <div>🔍</div> <div>📄</div> <div>🔗</div> <div>🔔</div> <div>&lt;1m ago</div> </div>		

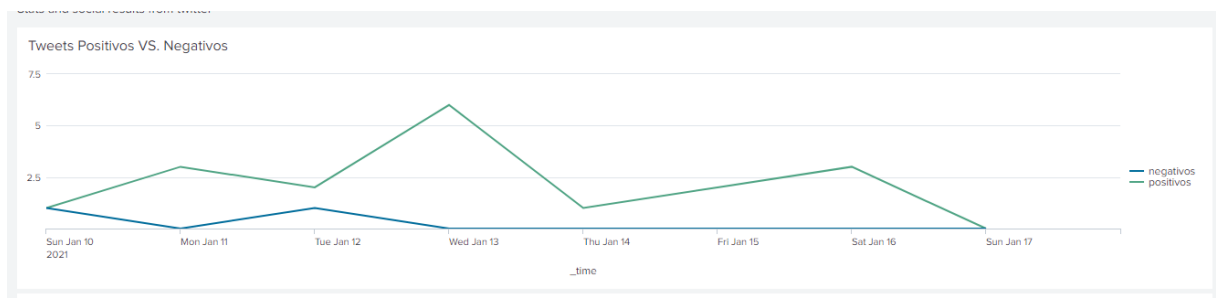
Fonte: Autor

**Figura 08 – Painéis com tweets positivos e negativos**

Tweets (possivelmente) positivos mais recentes			Tweets (possivelmente) negativos mais recentes		
Tweets positivos			tweets negativos		
_time ↕	tweet ↕	link ↕	_time ↕	tweet ↕	link ↕
2021-01-16 21:25:52	@HugoQueiroz_ @cafecomferri @AprendaComTC PARabens	https://twitter.com/luizthiagos/status/1350600188713443328	2021-01-12 22:35:43	@pedrohlyrab @vitorconor @AprendaComTC @cafecomferri @albuquerque_af @tatobarros @israel_massa Esse campeonato deveria ser cancelado. Burlar o game é foda e olhe q tô com 10% de valorização, mas operar delay é injusto. #cancelatorneioTC	https://twitter.com/_nilolobo/status/134916821
2021-01-16 20:26:40	@HugoQueiroz_ @cafecomferri @AprendaComTC Show!	https://twitter.com/juliovieiradc/status/1350585292256903169	2021-01-10 11:13:20	Essa bosta do @AprendaComTC mandando flash sms é das coisas mais irritantes do planeta terra. Tem como desabilitar essa desgraça? Porra @ClaroBrasil como	https://twitter.com/vitoraugusto/status/134827
2021-01-16 09:03:06	#IPO Equipe do TC Matrix está analisando os 43 prospectos na cvm para permitir que o investidor PF tenha a melhor informação sobre as ofertas e tome a melhor decisão de investimento. Acabei de postar a análise de Guararapes no @AprendaComTC. Convido todos para o trial de 7 dias.	https://twitter.com/HugoQueiroz_/status/1350585244295061505			
2021-01-15 16:02:44	@Tatobarros @AprendaComTC Muito feliz de fazer parte dessa equipe, brigada Tato!! Muitas novidades	https://twitter.com/brunabins_/status/1350585244295061505			

Fonte: Autor

**Figura 09 – Pannel com comparação cronológica de tweets positivos e negativos**

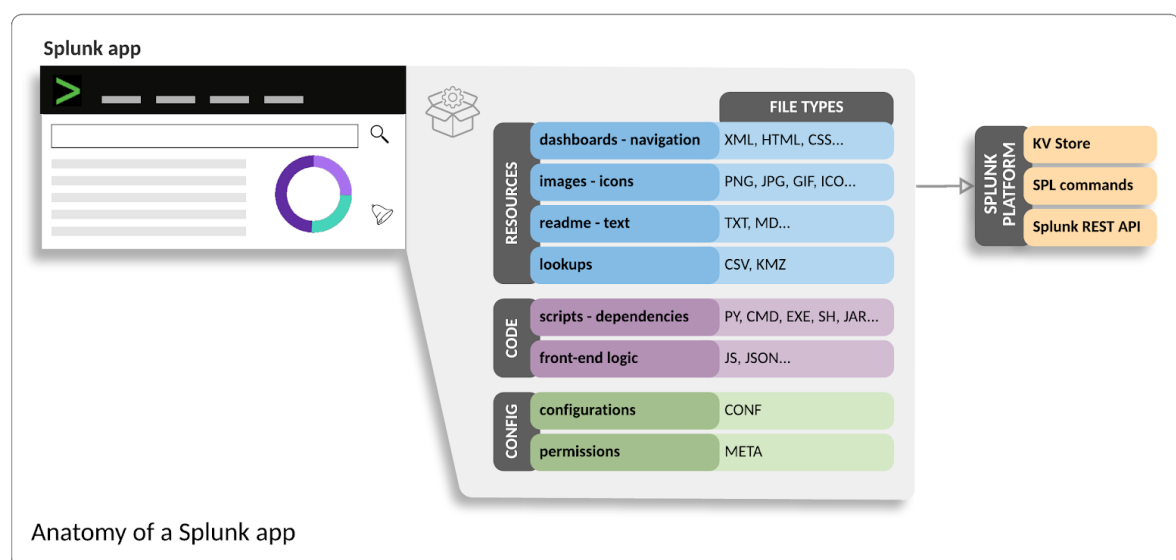


Fonte: Autor

Essas queries e dashboards podem ser colocadas dentro de um app, que é uma coleção de queries, dashboards, formatos de dados, imagens, alertas, scripts e outros recursos diversos. Em outras palavras, um app do Splunk é um conjunto de recursos necessários para uma ou mais funcionalidades.

O Splunk possui um procedimento simples para a criação, empacotamento, exportação e instalação de aplicativos - como a nossa solução, por exemplo. O desenvolvimento de aplicativos para Splunk é simples, mas extensível e potente. Isso permite que futuros contribuidores possam auxiliar o projeto com grande eficiência, e que usuários possam instalar a solução do zero ou atualizá-la com facilidade.

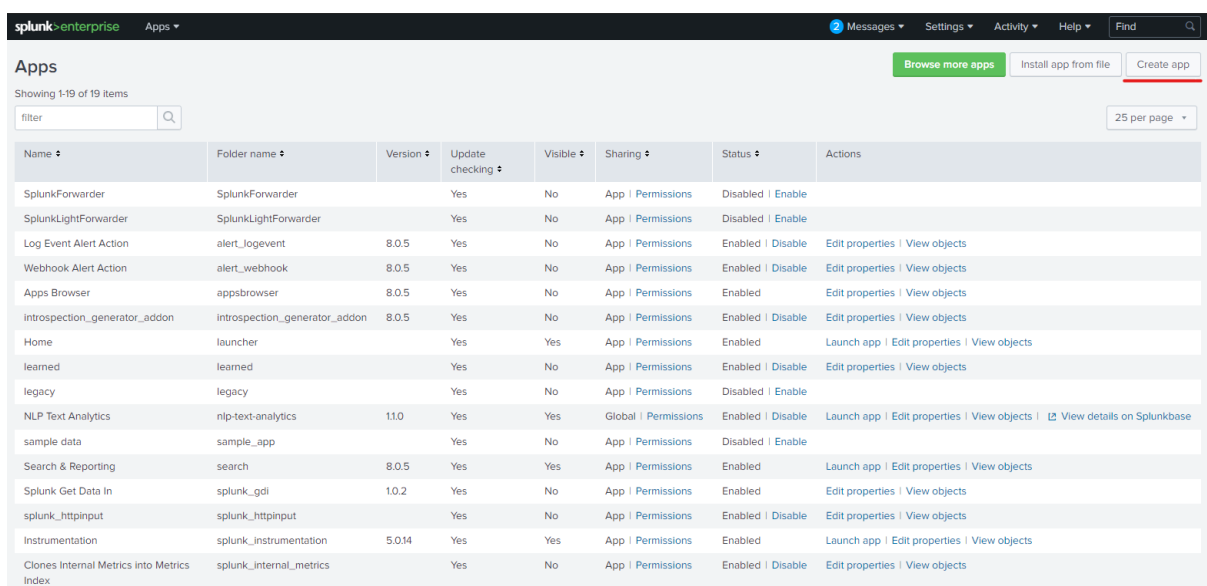
**Figura 10** – Estrutura de um aplicativo do Splunk



Fonte: <https://dev.splunk.com/enterprise/docs/developapps/appanatomy>. Acesso em 19 mar 2021

O usuário pode facilmente criar um app novo por um botão na interface gráfica do Splunk, como demonstrado na imagem abaixo.

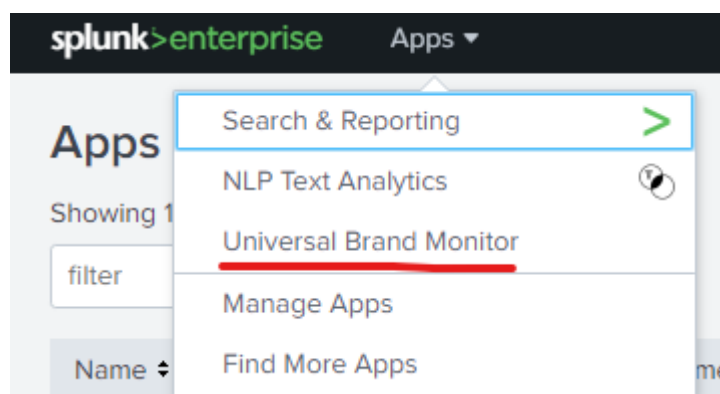
**Figura 11 – Botão para criação de aplicativo no Splunk**



Fonte: Autor

Isso cria uma seção especial para o app dentro da interface gráfica do Splunk, como demonstrado na imagem abaixo.

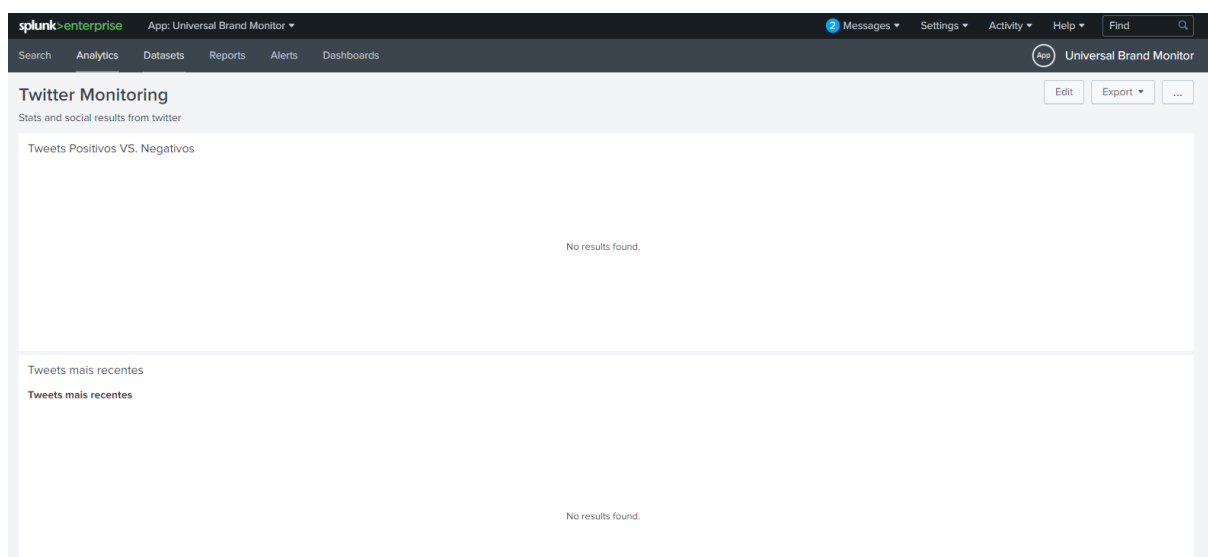
**Figura 12 – Aba de acesso à aplicativo do Splunk**



Fonte: Autor

O usuário pode em seguida criar uma Dashboard dentro do aplicativo, como demonstrado na imagem abaixo.

**Figura 13** – Dashboard em branco de aplicativo do Splunk



Fonte: Autor

Após o desenvolvimento do aplicativo, ele pode ser exportado ao executar o seguinte comando pela linha de comando: “splunk package app <nome do app>”

Isso cria um arquivo com a extensão “.spl”, que é apenas um formato compactado da pasta do aplicativo. A imagem abaixo mostra o resultado da execução do comando.

**Figura 14** – Comando para exportação de aplicativo do Splunk

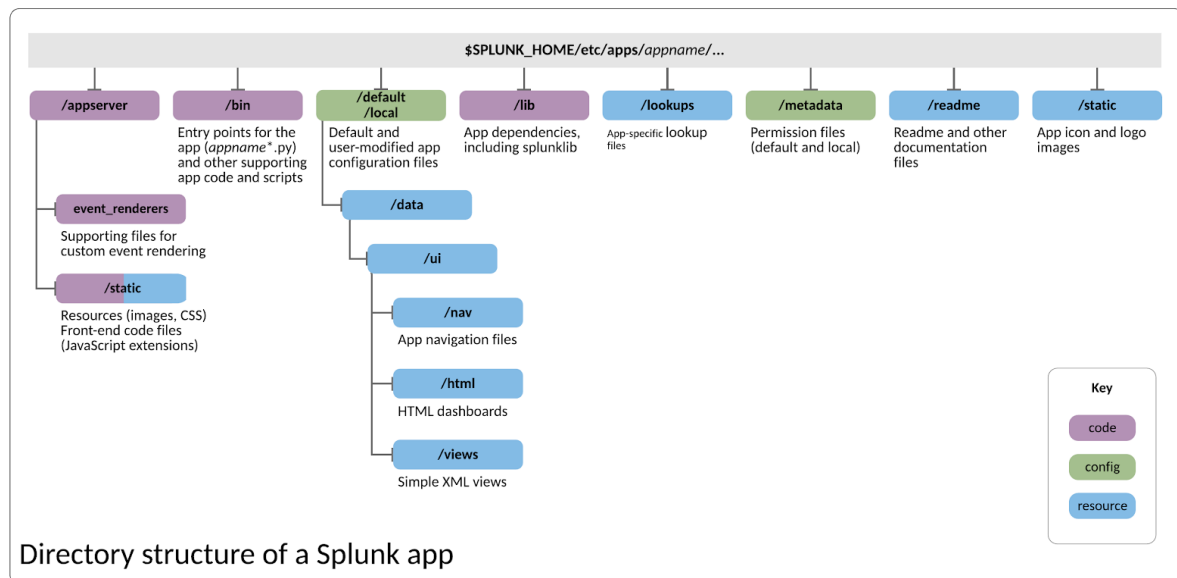
```
C:\Program Files\Splunk\bin>splunk package app ubm-app
Splunk username: admin
Password:
App 'ubm-app' is packaged.
Package location: C:\Program Files\Splunk\share\splunk\app_packages\ubm-app.spl
C:\Program Files\Splunk\bin>
```

Fonte: Autor

A pasta do aplicativo contém todos os recursos necessários para que o app execute de forma autocontida em outro servidor. Em outras palavras, todas as dependências do aplicativo são embutidas na pasta.

A imagem abaixo mostra a estrutura de diretórios padrão para um aplicativo do Splunk e a explicação, em inglês, da função de cada pasta.

**Figura 15** – Estrutura de diretórios de um aplicativo do Splunk



Fonte: Autor

Após gerar o arquivo “.spl” do aplicativo, basta que o desenvolvedor o distribua ou torne-o público. Desta forma, outros usuários poderão realizar o download deste arquivo “.spl” e facilmente instalar o aplicativo pela interface gráfica do Splunk. A imagem a seguir mostra o painel de instalação de aplicativos do Splunk. O usuário deve somente clicar em “Escolher arquivo” e realizar o upload do arquivo “.spl” para que o aplicativo seja instalado.

**Figura 16** – Menu para upload e instalação de aplicativo do Splunk

**Upload an app**

If you have a .spl or .tar.gz app file to install, you can upload it using this form.

You can replace an existing app via the Splunk CLI. [Learn more.](#)

File

Nenhum arquivo selecionado

☐ Upgrade app. Checking this will overwrite the app if it already exists.

Fonte: Autor

Outra opção de distribuição do aplicativo é pela loja de aplicativos oficial da Splunk, um repositório de aplicativos que pode ser acessado pela própria interface gráfica. No entanto, este processo requer uma análise oficial do aplicativo pela empresa responsável pelo Splunk e centraliza a distribuição do app em um único ponto. Isso vai, de certa forma, contra o resultado esperado - desejamos que desenvolvedores possam manter versões diferentes da aplicação com focos em certas redes sociais ou capacidades e, possivelmente, com configurações adicionais para melhor adaptá-la a certos ambientes ou propósitos.



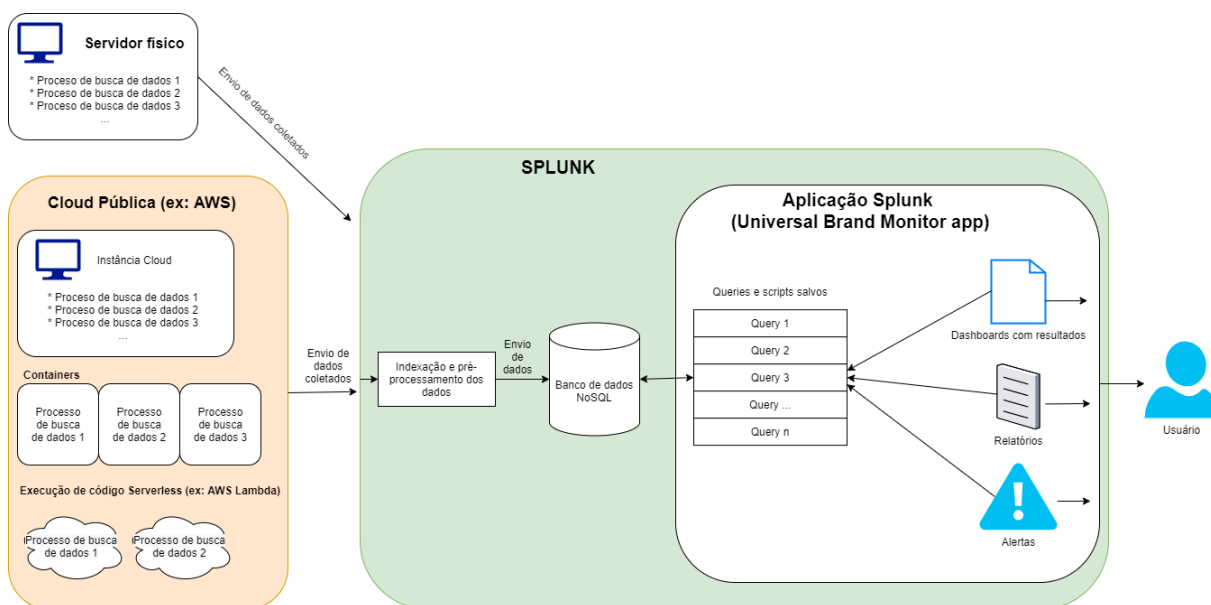
### **3 - SOLUÇÃO PROPOSTA: UNIVERSAL BRAND MONITOR (UBM)**

Nesta seção vamos descrever a solução construída, como ela funciona no geral e qual a sua arquitetura.

Apelidamos a solução de “Universal Brand Monitor”. Ela recebe esse nome pois deverá poder monitorar qualquer brand (“marca”, em português). Isso inclui, conforme dito anteriormente, desde alvos como empresas e pessoas até temas ou ideias. A partir deste ponto, utilizaremos a sigla “UBM” para designar a solução.

Um diagrama completo da arquitetura do UBM é exibido abaixo.

**Figura 17** – Arquitetura geral da solução (UBM)



Fonte: Autor

O UBM consiste em uma série de robôs para coleta de dados em arquitetura distribuída, um sistema central para o recebimento e filtragem dos dados, e finalmente um motor para a extração de informações úteis, alertas e inteligência a partir dos dados brutos coletados. Estes robôs para coleta de dados podem ser desenvolvidos em qualquer linguagem e buscar resultados em qualquer rede social - o único fator crucial é que eles enviem dados com um formato bem delimitado (como por exemplo em JSON ou XML) e não-duplicados para o Splunk pois o Splunk.

A delimitação é necessária para que o Splunk sempre saiba onde cada campo de informação começa e onde termina, permitindo que ele indexe os dados de forma satisfatória.

O envio de dados não-duplicados é necessário pois o Splunk não possui uma funcionalidade para descartar eventos duplicados na hora de ingestão dos dados. É possível realizar uma filtragem de eventos duplicados na hora de realizar buscas, mas os eventos duplicados acabam ocupando um espaço desnecessário em disco e requerem que todas as queries sejam reescritas para explicitamente filtrar eventos duplicados.

No diagrama exemplificamos dois ambientes de execução possíveis para os robôs coletores do UBM: Servidores físicos (também conhecidos como servidores on-premises) e Clouds Públicas como Amazon AWS, Microsoft Azure e Google Cloud Platform em diversas modalidades: dentro de instâncias (máquinas virtuais hospedadas pelo fornecedor de Cloud), dentro de containers (execução de containers Kubernetes, Docker ou tecnologias similares) ou em serviços serverless (onde o código é executado pelo provedor de Cloud sem um servidor fixo associado).

Para a coleta de dados do UBM, criamos um robô em Python 3 especializado na coleta de informações pelo Twitter. No entanto, poderíamos ter escolhido qualquer outra linguagem de programação ou alvo de monitoração.

O Twitter foi escolhido por ser uma rede social de enorme utilização, tanto por indivíduos quanto por empresas, e onde diversos conflitos ocorrem diariamente. Além disso, também possui um formato de postagem simples e, usando a biblioteca twint, que é open-source e gratuita, pode-se facilmente buscar por tweets contendo termos específicos ou que mencionem algum usuário, marca, empresa ou pessoa. A imagem abaixo mostra diversos exemplos de utilização do twint.

**Figura 18** – Exemplos de comandos disponíveis da biblioteca twint

## Basic Examples and Combos.

A few simple examples to help you understand the basics:

- `twint -u username` - Scrape all the Tweets from *user's* timeline.
- `twint -u username -s pineapple` - Scrape all Tweets from the *user's* timeline containing *pineapple*.
- `twint -s pineapple` - Collect every Tweet containing *pineapple* from everyone's Tweets.
- `twint -u username --year 2014` - Collect Tweets that were tweeted **before** 2014.
- `twint -u username --since 2015-12-20` - Collect Tweets that were tweeted since 2015-12-20.
- `twint -u username -o file.txt` - Scrape Tweets and save to file.txt.
- `twint -u username -o file.csv --csv` - Scrape Tweets and save as a csv file.
- `twint -g="48.880048,2.385939,1km" -o file.csv --csv` - Scrape Tweets from a radius of 1km around a place in Paris and export them to a csv file.
- `twint -u username -es localhost:9200` - Output Tweets to Elasticsearch
- `twint -u username -o file.json --json` - Scrape Tweets and save as a json file.
- `twint -u username --database tweets.db` - Save Tweets to a SQLite database.
- `twint -u username --followers` - Scrape a Twitter user's followers.
- `twint -u username --following` - Scrape who a Twitter user follows.
- `twint -u username --favorites` - Collect all the Tweets a user has favorited.

Fonte: <https://github.com/twintproject/twint/wiki/Basic-usage>. Acesso em 18 mar. 2021

O twint retorna os resultados num formato de “melhor esforço”: não há garantia que todos os resultados serão encontrados, que eles virão em uma ordem específica ou que não existirão resultados duplicados.

O robô de monitoração do Twitter usa o twint para realizar uma busca por termos pré-definidos pelo usuário e coleta todos os tweets que contém o termo. Cada tweet possui um número identificador único, abreviado de id. Esse id é inserido em um banco de dados local leve (sqlite3) para realizar uma pré-filtragem dos dados, evitando a coleta de tweets duplicados. Os dados então são enviados para o Splunk usando a biblioteca oficial do Splunk para Python chamada de “Splunk Enterprise SDK” [9]

Observe que, numa arquitetura distribuída, um usuário poderia programar a execução de diversos robôs buscando por termos diferentes em redes sociais diferentes.

Seguindo o fluxo dos dados, a próxima parte da arquitetura do UBM é o Splunk. Ele é o responsável por todo o recebimento e processamento dos dados, pela interface do usuário e pela geração de resultados e relatórios. Os robôs coletores de dados do UBM enviam os dados para o Splunk por uma das diversas formas disponíveis. No caso utilizamos uma biblioteca para Python, mas outros robôs poderiam utilizar outros (ou os mesmos) métodos de forma paralela. Ao receber os dados, o Splunk realiza o pré-processamento e indexação dos mesmos. Essas operações são nativas do Splunk e completamente transparentes tanto para o usuário quanto

para desenvolvedores, então não vamos entrar em detalhes aqui sobre como funcionam - apenas vale mencionar que esse processo é crucial para a agilidade de execução das buscas do Splunk e o bom gerenciamento de seus recursos.

Em seguida, os dados têm seus campos extraídos: o texto bruto que chega no Splunk é dividido em setores ou áreas relevantes para facilitar a busca no futuro. A imagem a seguir mostra o texto JSON bruto na parte superior e os campos extraídos desse texto na parte inferior.

**Figura 19** – Comparação entre dados recebidos e campos extraídos pelo Splunk

i

Time

Event

timestamp: 10:28:25

timezone: -0300

trans\_dest:

trans\_src:

translate:

tweet: @LucasCostaAT @AprendaComTC Lucas, dá uma olhada nas trilhas de conhecimento, tá muito legal

urls: [ [+]

]

user\_id: 208244934

user\_id\_str: 208244934

user\_rt:

user\_rt\_id:

username: freazav

video: 0

}

Show as raw text

Event Actions

Type	<input checked="" type="checkbox"/> Field	Value	Actions
Selected	<input checked="" type="checkbox"/> host	127.0.0.1	
	<input checked="" type="checkbox"/> source	http-simple	
	<input checked="" type="checkbox"/> sourcetype	twitter	
Event	<input type="checkbox"/> conversation_id	1369269146270175236	
	<input type="checkbox"/> datefull	2021-03-09 10:28:25	
	<input type="checkbox"/> datestamp	2021-03-09	
	<input type="checkbox"/> datetime	2021-03-09 10:28:25 E. South America Standard Time	
	<input type="checkbox"/> geo		
	<input type="checkbox"/> id	1369278903169212418	
	<input type="checkbox"/> id_str	1369278903169212418	

...-24h%40b&latest=now&q=search index%3

Fonte: Autor

Para esta instalação do UBM usamos apenas o aplicativo do Splunk rodando em uma máquina pessoal, mas conforme explicado anteriormente seria possível utilizar diferentes arquiteturas do Splunk dependendo do volume esperado pelo UBM e das necessidades do usuário.

Em seguida foram criadas as queries - pequenos scripts na linguagem própria do Splunk (SPL) para a busca e processamento dos dados captado que podem formar gráficos, tabelas e outras formas de visualização de resultados.

Abaixo exibimos um exemplo de query utilizada para filtrar por tweets possivelmente positivos. Para demonstrar melhor a sintaxe e funcionamento da SPL, também iremos comentar o propósito de cada linha.

**Figura 20** – Query para filtrar tweets positivos



_time	tweet	link
2021-03-09 10:28:25	@LucasCostaAT @AprendaComTC Lucas, dá uma olhada nas trilhas de conhecimento, tá muito legal	<a href="https://twitter.com/freazav/status/1369278903169212418">https://twitter.com/freazav/status/1369278903169212418</a>
2021-03-08 20:03:03	"Isolamento e quarentena são fundamentais para frear a transmissão do #COVID19; período requer bom senso, paciência e uma boa dose de criatividade." - Newsletter Traders Club 23/03/2020. @AprendaComTC	<a href="https://twitter.com/donmlander/status/1369061127003512834">https://twitter.com/donmlander/status/1369061127003512834</a>
2021-03-08 19:58:41	Depois do Matrix, o Plus é a melhor sacada do @AprendaComTC até hoje...	<a href="https://twitter.com/torresac/status/1369060028175499276">https://twitter.com/torresac/status/1369060028175499276</a>
2021-03-05 21:11:33	@AprendaComTC [=] com o que a plataforma tem oferecido. São constantes as atualizações e os conteúdos parecem se multiplicar a cada dia! Estou muito feliz com o aprendizado e tenho certeza absoluta que os resultados virão em breve!	<a href="https://twitter.com/javepaulo/status/1367991204625932289">https://twitter.com/javepaulo/status/1367991204625932289</a>
2021-03-05 16:50:43	@wilson_ramires @AprendaComTC cada dia melhor!!!	<a href="https://twitter.com/romanicero2/status/1367925561952190466">https://twitter.com/romanicero2/status/1367925561952190466</a>
2021-03-05 08:57:45	Albuquerque e equipe do @AprendaComTC sou testemunha da evolução constante e intensa do TC desde quando iniciei, logo no início. Um trabalho realmente admirável e incansável na busca de fazer o melhor ainda melhor.	<a href="https://twitter.com/MilitaoAbantes/status/1367806534747557889">https://twitter.com/MilitaoAbantes/status/1367806534747557889</a>
2021-03-03 12:05:25	@PontesRetorno @AprendaComTC Muito bom, Pontes!	<a href="https://twitter.com/mariodeavelar/status/1367128989366689796">https://twitter.com/mariodeavelar/status/1367128989366689796</a>

Fonte: Autor

Na primeira linha, estamos declarando que os dados devem ser buscados do “index=main”, que é o index principal do Splunk, e que o “sourcetype” dos dados deve ser igual à “twitter”. Esse sourcetype é declarado pelo robô coletor do UBM quando envia os dados e serve para organizar os dados de origens diferentes.

Na segunda linha, a função “eval” é usada para calcular um campo novo, resultante da junção da data com o horário do tweet para criar o horário completo de postagem.

Na terceira linha, a função “eval” é usada novamente junto com a função “strptime” para transformar a data completa, que está em formato texto, para um formato epoch. O resultado desse cálculo é armazenado na coluna (ou variável) “\_time”, que é a coluna utilizada pelo Splunk para gerar estatísticas e filtros baseados em datas.

Na quarta linha, a função “sort” é usada para ordenar os eventos por ordem cronológica descendente.

Na quinta linha, a função “where” é usada para realizar uma segunda busca nos dados processados até o momento, mantendo apenas os resultados com menos de 7 dias de diferença para o momento atual. Em outras palavras, resultados com mais de 7 dias de idade são descartados.

Na sexta linha, usamos o comando “search” para buscar por termos específicos dentro dos resultados. Qualquer evento que não contenha pelo menos uma das palavras listadas será descartado.

Na sétima e última linha usamos o comando “table” para exibir os eventos restantes em uma tabela contendo as colunas “\_time” (horário do tweet), “tweet” (conteúdo do tweet) e “link” (URL completa do tweet).

Uma query é apenas uma busca ou um processamento de dados específicos, mas ela pode ser utilizada para alguns propósitos. Um deles é transformar a query em um alerta, fazendo com que ela seja executada em uma frequência determinada e gere um alerta (via e-mail, SMS, requisição HTTP ou qualquer outro método) caso alguma condição lógica se concretize.

Outro uso comum para queries é aglutinar diversas delas em um único dashboard, como já mencionado na seção de revisão bibliográfica.

O UBM dispõe de uma dashboard com diversas queries retornando resultados, estatísticas e fatos relevantes sobre a presença do(s) alvo(s) no Twitter. Tais resultados são baseados nos tweets coletados. Essa dashboard será exibida no próximo capítulo.

Também foi criado um manual de instalação completo para o UBM. Tal manual é curto, tendo apenas 4 páginas, e a instalação completa do UBM pode ser feita em poucos passos. Estimamos que a primeira instalação da aplicação feita por um usuário inexperiente leve menos de uma hora.

O aplicativo do UBM foi empacotado e disponibilizado em um repositório no Github<sup>[10]</sup>, marcando a primeira release pública do projeto.

#### **4 – TESTES E ANÁLISE DE RESULTADOS**

Como dito anteriormente, criamos robôs dentro do UBM para a monitoração da rede Twitter. Nesta seção vamos explorar como os dados foram coletados, tratados e processados pelo UBM, e quais resultados foram gerados.

A coleta dos dados foi feita por meio do robô coletor do Twitter, já citado anteriormente. Ao iniciar o robô ele verifica se já há um banco de dados SQLite local. Se não existe, ele cria um. O módulo twint também é configurado para realizar uma busca mais precisa:

**Figura 21** – Configurações utilizadas para a biblioteca twint

```
# Configure
c = twint.Config()
c.Store_object = True
c.Store_object_tweets_list = resultados
c.Hide_output = True
c.Limit = tweetsParaBuscar
c.Store_json = True
c.Filter_retweets = False
c.Profile_fulle = True
```

Fonte: Autor

Neste caso buscamos apenas por tweets citando ou contendo uma conta de uma organização, @AprendaComTC<sup>[11]</sup>:

**Figura 22** – Definição de termos a serem buscados

```
resultadosTotal = []
resultadosTotal = resultadosTotal + procurarTweets("AprendaComTC")
```

Fonte: Autor

Vale deixar claro que poderíamos pesquisar também por outras contas na mesma busca, buscar apenas por replies à uma ou mais contas, buscar apenas tweets de certas pessoas ou quaisquer outras diferentes. Escolhemos esse método por ser o mais abrangente e pegar todas as menções à conta.

A busca do twint é realizada e os dados são recebidos inicialmente como objetos dentro do Python. Cada atributo do objeto possui uma informação do tweet coletado (o identificador único, o conteúdo, o número de likes, o horário que foi postado, etc).

**Figura 23** – Visão de dados brutos recuperados pelo Twint

```
H:\GoogleDrive\Pessoal\TCC\universalbrandmonitor>python ColetorTwitter.py
Banco de dados ja existe
AprendaComTC - 600 tweets coletados
{"id": 1371188532035923970, "id_str": "1371188532035923970", "conversation_id": "1371113638954668032", "datetime": "2021-03-14 16:56:36 E. South A
merica Standard Time", "datestamp": "2021-03-14", "timestamp": "16:56:36", "user_id": 756318290, "user_id_str": "756318290", "username": "zeelias
almeida", "name": "Jose Elias Almeida", "place": "", "timezone": "-0300", "mentions": [], "reply_to": [{"screen_name": "AprendaComTC", "name": "TC
- TradersClub", "id": "767763810409181188"}], "urls": [], "photos": [], "video": 0, "thumbnail": "", "tweet": "@AprendaComTC \ud83d\udc4f\ud83d\u
dc4f\ud83d\udc4f\ud83d\udc4f\ud83d\udc4f", "lang": "und", "hashtags": [], "cashtags": [], "replies_count": 0, "retweets_count": 0, "likes_count":
1, "link": "https://twitter.com/zeelias_almeida/status/1371188532035923970", "retweet": false, "retweet_id": "", "retweet_date": "", "user_rt": "",
"user_rt_id": "", "quote_url": "", "near": "", "geo": "", "source": "", "translate": "", "trans_src": "", "trans_dest": ""}

{"id": 1371174588495892482, "id_str": "1371174588495892482", "conversation_id": "1371156126033711104", "datetime": "2021-03-14 16:01:11 E. South A
merica Standard Time", "datestamp": "2021-03-14", "timestamp": "16:01:11", "user_id": 2255272568, "user_id_str": "2255272568", "username": "leolev
atti", "name": "Leonardo Levatti", "place": "", "timezone": "-0300", "mentions": [], "reply_to": [{"screen_name": "Boghosian", "name": "Boghosian",
"id": "39934516"}, {"screen_name": "jlbbraga", "name": "JL Braga, CFA", "id": "24367340"}, {"screen_name": "AprendaComTC", "name": "TC - TradersC
lub", "id": "767763810409181188"}], "urls": [], "photos": [], "video": 0, "thumbnail": "", "tweet": "@Boghosian @jlbbraga @AprendaComTC Som no caix
al! Rsl! \ud83d\udc4f\ud83d\udc4f\ud83d\udc4f", "lang": "pt", "hashtags": [], "cashtags": [], "replies_count": 0, "retweets_count": 0, "likes_count": 1, "link
": "https://twitter.com/leolevatti/status/1371174588495892482", "retweet": false, "retweet_id": "", "retweet_date": "", "user_rt": "", "user_rt_id
": "", "quote_url": "", "near": "", "geo": "", "source": "", "translate": "", "trans_src": "", "trans_dest": ""}

{"id": 1371171417258143744, "id_str": "1371171417258143744", "conversation_id": "1371156126033711104", "datetime": "2021-03-14 15:48:35 E. South A
merica Standard Time", "datestamp": "2021-03-14", "timestamp": "15:48:35", "user_id": 39934516, "user_id_str": "39934516", "username": "boghosian",
"name": "Boghosian", "place": "", "timezone": "-0300", "mentions": [{"screen_name": "leolevatti", "name": "leonardo levatti", "id": "2255272568"}],
{"screen_name": "aprendacomtc", "name": "tc - tradersclub", "id": "767763810409181188"}], "reply_to": [{"screen_name": "jlbbraga", "name": "JL B
raga, CFA", "id": "24367340"}], "urls": [], "photos": [], "video": 0, "thumbnail": "", "tweet": "@jlbbraga Vamos fazer um programa , \u201cplaylist
dos gestores\u201d na TC radio ? @leolevatti @AprendaComTC", "lang": "pt", "hashtags": [], "cashtags": [], "replies_count": 1, "retweets_count":
0, "likes_count": 8, "link": "https://twitter.com/Boghosian/status/1371171417258143744", "retweet": false, "retweet_id": "", "retweet_date": "", "
user_rt": "", "user_rt_id": "", "quote_url": "", "near": "", "geo": "", "source": "", "translate": "", "trans_src": "", "trans_dest": ""}
```

Fonte: Autor

Cada tweet possui um id único. O robô verifica no banco de dados se esse id já foi visto antes - caso positivo, o tweet é descartado.

**Figura 24** – Código responsável pela remoção de resultados duplicados

```
def removerDuplicados(resultados):
    resultadosNovos = []
    for result in resultados:
        tweetid = result.id
        if(not verificarTweetJaVisto(tweetid)):
            resultadosNovos.append(result)
    return resultadosNovos
```

Fonte: Autor

**Figura 25** – Código responsável pela verificação de resultados duplicados

```
def verificarTweetJaVisto(tweetId):
    with dbConnection:
        data = dbConnection.execute("SELECT COUNT(*) FROM TWEETS WHERE tweetid = " + str(tweetId))
        for row in data:
            if (row[0] == 0):
                return False
            else:
                return True
```

Fonte: Autor

Os tweets novos são transformados em um JSON onde cada atributo do objeto é um campo diferente, e então são enviados para o Splunk no sourcetype “twitter” (isso é feito apenas por uma questão de organização e não depende que o usuário crie essa sourcetype anteriormente, ele é automaticamente criado):



**Figura 26** – Código responsável pela transformação de dados brutos em JSON

```
def createTweetJson(tweet):
    tweetDict = vars(tweet)
    tweetJson = json.dumps(tweetDict)
    return tweetJson
```

Fonte: Autor

**Figura 27** – Código responsável pela conexão e envio de dados ao Splunk

```
def openSplunkConnection():
    service = client.connect(host='localhost',port=8089,username=splunkUsername,password=splunkPassword)
    indexConnection = service.indexes["main"]
    return indexConnection

def sendJsonToSplunk(jsonData, indexConnection):
    indexConnection.submit(jsonData, sourcetype="twitter")
```

Fonte: Autor

Os tweets que foram enviados também são adicionados ao banco para evitar que se enviem duplicatas deles no futuro:

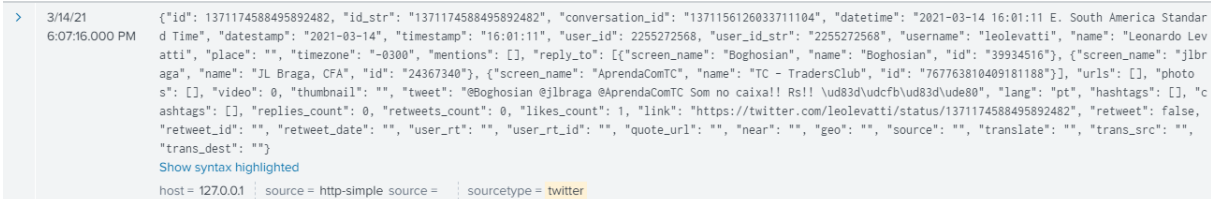
**Figura 28** – Código responsável pela adição de resultados à lista de resultados já vistos

```
def adicionarATweetsJavistos(tweet):
    try:
        with dbConnection:
            dbConnection.execute(" INSERT INTO TWEETS VALUES (" + str(tweet.id) + ")")
    except:
        pass
```

Fonte: Autor

Conforme citado anteriormente, os tweets que chegam ao Splunk passam por uma indexação e extração de campos do próprio Splunk. O tweet chega da seguinte forma para o Splunk, como um bloco de texto:

**Figura 29** – Dados brutos recebidos pelo Splunk



The screenshot shows a Splunk search interface. On the left, there's a search bar with the text "> 3/14/21 6:07:16.000 PM". The main area displays a single search result as a JSON object. The JSON contains fields like 'id', 'id\_str', 'conversation\_id', 'datetime', 'timestamp', 'user\_id', 'user\_id\_str', 'username', 'name', 'place', 'timezone', 'mentions', 'reply\_to', 'screen\_name', 'name', 'id', 'photo', 'video', 'thumbnail', 'tweet', 'lang', 'hashtags', 'replies\_count', 'retweets\_count', 'likes\_count', 'link', 'retweet\_id', 'retweet\_date', 'user\_rt', 'user\_rt\_id', 'quote\_url', 'near', 'geo', 'source', 'translate', and 'trans\_src'. At the bottom, there's a 'Show syntax highlighted' link and a status bar showing 'host = 127.0.0.1 | source = http-simple source = | sourcetype = twitter'.

Fonte: Autor

O Splunk processa esse bloco de texto para que o tweet adquira esse formato, com os campos identificados para facilitar a visualização:

**Figura 30** – Dado organizado gerado pelo Splunk

```
> 3/14/21 6:07:16.000 PM { [-]
  cashtags: [ [+]
  ]
  conversation_id: 1371156126033711104
  datestamp: 2021-03-14
  datetime: 2021-03-14 16:01:11 E. South America Standard Time
  geo:
  hashtags: [ [+]
  ]
  id: 1371174588495892500
  id_str: 1371174588495892482
  lang: pt
  likes_count: 1
  link: https://twitter.com/leolevatti/status/1371174588495892482
  mentions: [ [+]
  ]
  name: Leonardo Levatti
  near:
  photos: [ [+]
  ]
  place:
  quote_url:
  replies_count: 0
  reply_to: [ [+]
  ]
  retweet: false
  retweet_date:
  retweet_id:
  retweets_count: 0
  source:
  thumbnail:
  timestamp: 16:01:11
  timezone: -0300
  trans_dest:
  trans_src:
  translate:
  tweet: @Boghossian @j1braga @AprendaComTC Som no caixa!! Rs!! 📺🔗
  urls: [ [+]
  ]
  user_id: 2255272568
  user_id_str: 2255272568
  user_rt:
  user_rt_id:
  username: leolevatti
  video: 0
}
Show as raw text
host = 127.0.0.1 | source = http-simple source = | sourcetype = twitter
```

Fonte: Autor

Os campos interessantes também são extraídos e indexados para facilitar e agilizar buscas envolvendo esses campos:

**Figura 31** – Campos dentro do dado extraídos pelo Splunk

Type	<input checked="" type="checkbox"/> Field	Value	Actions
Selected	<input checked="" type="checkbox"/> host ▾	127.0.0.1	▼
	<input checked="" type="checkbox"/> source ▾	http-simple	▼
			▼
	<input checked="" type="checkbox"/> sourcetype ▾	twitter	▼
Event	<input type="checkbox"/> conversation_id ▾	1371156126033711104	▼
	<input type="checkbox"/> datestamp ▾	2021-03-14	▼
	<input type="checkbox"/> datetime ▾	2021-03-14 16:01:11 E. South America Standard Time	▼
	<input type="checkbox"/> geo ▾		▼
	<input type="checkbox"/> id ▾	1371174588495892482	▼
	<input type="checkbox"/> id_str ▾	1371174588495892482	▼
	<input type="checkbox"/> lang ▾	pt	▼
	<input type="checkbox"/> likes_count ▾	1	▼
	<input type="checkbox"/> link ▾	https://twitter.com/leolevatti/status/1371174588495892482	▼
	<input type="checkbox"/> name ▾	Leonardo Levatti	▼
	<input type="checkbox"/> near ▾		▼
	<input type="checkbox"/> place ▾		▼
	<input type="checkbox"/> quote_url ▾		▼
	<input type="checkbox"/> replies_count ▾	0	▼
	<input type="checkbox"/> reply_to().id ▾	39934516	▼
		24367340	▼
		767763810409181188	▼
	<input type="checkbox"/> reply_to().name ▾	Boghosian	▼
		JL Braga, CFA	▼
		TC - TradersClub	▼
	<input type="checkbox"/> reply_to().screen_name ▾	Boghosian	▼
		jbraga	▼
		AprendaComTC	▼
	<input type="checkbox"/> retweet ▾	false	▼
	<input type="checkbox"/> retweet_date ▾		▼
	<input type="checkbox"/> retweet_id ▾		▼
	<input type="checkbox"/> retweets_count ▾	0	▼
	<input type="checkbox"/> thumbnail ▾		▼
	<input type="checkbox"/> timestamp ▾	none	▼
		16.01.11	▼
	<input type="checkbox"/> timezone ▾	-0300	▼
	<input type="checkbox"/> trans_dest ▾		▼
	<input type="checkbox"/> trans_src ▾		▼
	<input type="checkbox"/> translate ▾		▼
	<input type="checkbox"/> tweet ▾	@Boghosian @jbraga @AprendaComTC Sem no calca! Rsl! 🚗🚗	▼
	<input type="checkbox"/> user_id ▾	2255272568	▼
	<input type="checkbox"/> user_id_str ▾	2255272568	▼
	<input type="checkbox"/> user_rt ▾		▼
	<input type="checkbox"/> user_rt_id ▾		▼
	<input type="checkbox"/> username ▾	leolevatti	▼
<input type="checkbox"/> video ▾	0	▼	
Time	_time ▾	2021-03-14T18:07:16.000-03:00	
Default	<input type="checkbox"/> index ▾	main	▼
	<input type="checkbox"/> linecount ▾	1	▼
	<input type="checkbox"/> punct ▾	[{"ts": 1615785671.0, "type": "tweet", "source": "twitter", "target": "leolevatti", "value": {"conversation_id": 1371156126033711104, "datestamp": "2021-03-14", "datetime": "2021-03-14 16:01:11 E. South America Standard Time", "geo": null, "id": 1371174588495892482, "id_str": "1371174588495892482", "lang": "pt", "likes_count": 1, "link": "https://twitter.com/leolevatti/status/1371174588495892482", "name": "Leonardo Levatti", "near": null, "place": null, "quote_url": null, "replies_count": 0, "reply_to": {"id": 39934516, "id_str": "24367340", "name": "Boghosian", "screen_name": "jbraga", "username": "AprendaComTC"}, "retweet": false, "retweet_date": null, "retweet_id": null, "retweets_count": 0, "thumbnail": null, "timestamp": "none", "timezone": "-0300", "trans_dest": null, "trans_src": null, "translate": null, "tweet": "@Boghosian @jbraga @AprendaComTC Sem no calca! Rsl! 🚗🚗", "user_id": 2255272568, "user_id_str": "2255272568", "user_rt": null, "user_rt_id": null, "username": "leolevatti", "video": 0}}]	▼
	<input type="checkbox"/> splunk_server ▾	DESKTOP-SSDH4PT	▼

Fonte: Autor

Um detalhe interessante que vale ser mencionado é que os coletores não precisam estar rodando continuamente - sempre ao executar o coletor ele irá automaticamente coletar

---

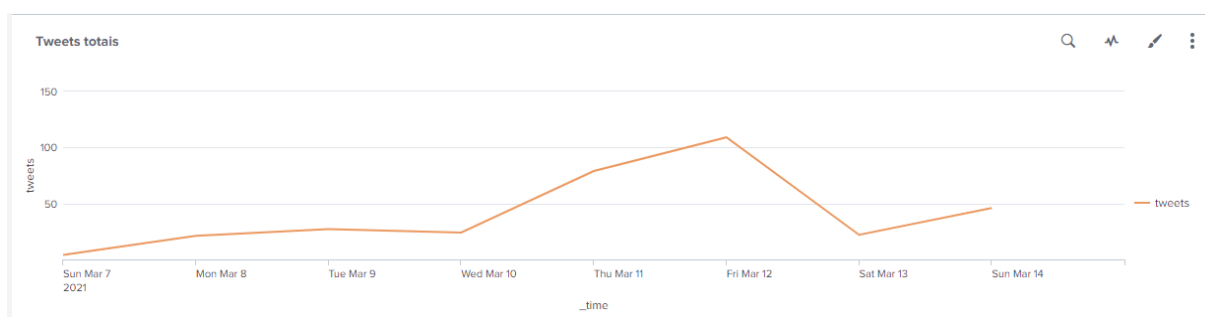
UFABC – Projeto de Graduação em Computação III 27 | P á g i n a

todos os novos tweets desde a última execução. Não há um período máximo entre execuções onde dados começam a ser perdidos. A única exceção seria devido aos tweets deletados, que obviamente não podem ser coletados após a sua deleção.

Como já citado anteriormente, os dados são agregados em dashboards para a visualização do usuário. Vamos agora estudar os dados apresentados pelos dashboards do UBM e quais foram os resultados obtidos.

A figura abaixo mostra a contagem por dia de tweets mencionando a marca nos últimos 7 dias. Conseguimos perceber altas e baixas de interação com a conta.

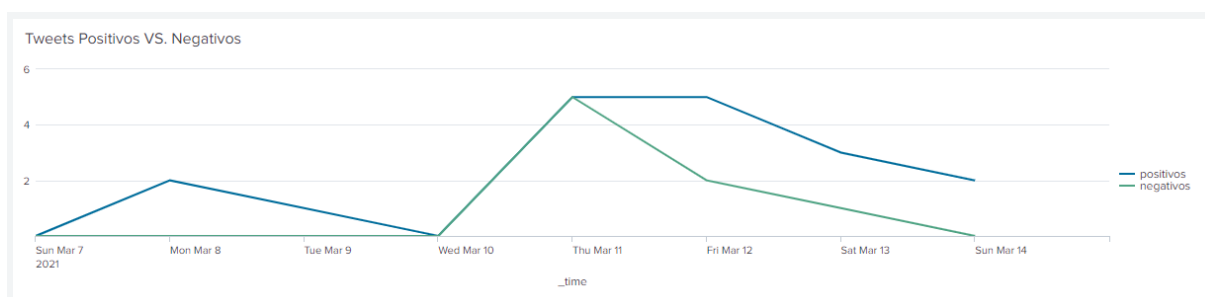
**Figura 32** – Painel com menções diárias à marca



Fonte: Autor

A figura abaixo mostra a contagem por dia de tweets negativos comparados com tweets positivos nos últimos 7 dias. Vamos detalhar melhor como um tweet é considerado positivo ou negativo em breve.

**Figura 33** – Painel com menções diárias positivas e negativas à marca



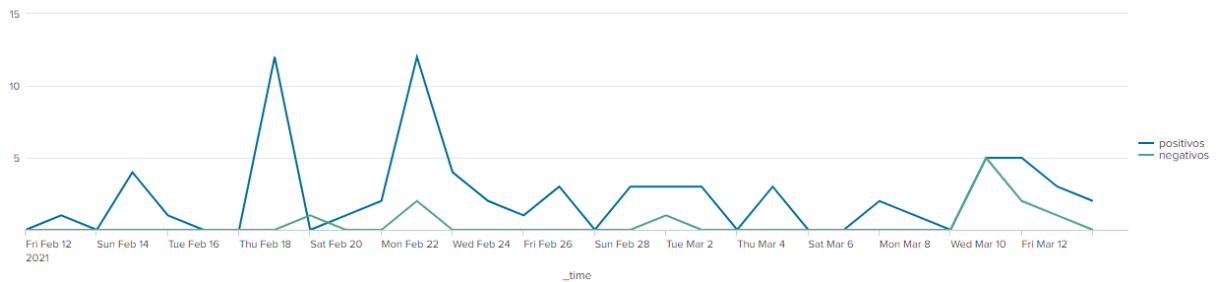
Fonte: Autor

O gráfico é interessante - imediatamente podemos perceber que houve um pico de tweets negativos no dia 11/03/2021. Num cenário de aplicação prática, seria interessante para o usuário analisar o conteúdo desses tweets para saber se eles possuem alguma relação entre si

ou se o volume elevado foi apenas uma coincidência - será que diversos usuários estão insatisfeitos com algum ponto específico? Se sim, como podemos ajudar esses usuários?

A imagem abaixo mostra a mesma visão anterior, mas num período de 30 dias ao invés de 7:

**Figura 34** – Painel com menções diárias positivas e negativas à marca em 30 dias



Fonte: Autor

Podemos perceber que alguns dias tiveram atividades intensas de tweets positivos (19/02/2021 e 23/02/2021), e que os tweets negativos são menos frequentes e menos intensos. Isso pode ser considerado um sinal de saúde da imagem pública da marca. Se o usuário percebesse que existe um volume muito maior de tweets negativos do que positivos, isso poderia ser considerado como um sinal de má reputação da marca, de problemas recorrentes ou de insatisfação geral.

A imagem abaixo mostra um painel dentro da dashboard que exhibe os tweets mais recentes (em ordem cronológica, não de captura):

**Figura 35 – Pannel com tweets mais recentes mencionando a marca**

Tweets mais recentes		
_time ↕	tweet ↕	link ↕
2021-03-14 18:30:12	Analisando Guararapes #GUAR3 e vendo o que temos para essa semana no TC Matrix. Enquanto isso, ouvindo aquela rádio na TC STATION do meu TC @AprendaComTC que é show de bola. Daqui a pouco isso vira um SuperApp. Dá pra ficar rapidamente bem informado e recebendo dicas diárias.	<a href="https://twitter.com/xpluis/status/1371212090606780417">https://twitter.com/xpluis/status/1371212090606780417</a>
2021-03-14 16:56:36	@AprendaComTC 🍌🍌🍌🍌🍌	<a href="https://twitter.com/zeelias_almeida/status/1371188532035923970">https://twitter.com/zeelias_almeida/status/1371188532035923970</a>
2021-03-14 16:01:11	@Boghosian @jlbraga @AprendaComTC Som no caixa!! Rs!! 📺📺	<a href="https://twitter.com/leolevatti/status/1371174588495892482">https://twitter.com/leolevatti/status/1371174588495892482</a>
2021-03-14 15:48:35	@jlbraga Vamos fazer um programa , "playlist dos gestores" na TC radio ? @leolevatti @AprendaComTC	<a href="https://twitter.com/Boghosian/status/1371171417258143744">https://twitter.com/Boghosian/status/1371171417258143744</a>
2021-03-14 15:21:19	@WZDondoni @guizilse @albuquerque_af @cafeacomferri @memolitinho @AprendaComTC Vamos providenciar	<a href="https://twitter.com/j_ramacciotti/status/1371164554386939904">https://twitter.com/j_ramacciotti/status/1371164554386939904</a>
2021-03-14 15:20:34	@guizilse @albuquerque_af @cafeacomferri @memolitinho @AprendaComTC Tmj Gui !!	<a href="https://twitter.com/j_ramacciotti/status/1371164367979479041">https://twitter.com/j_ramacciotti/status/1371164367979479041</a>
2021-03-14 14:48:07	@alaramaximo @Boghosian @AprendaComTC Acho uma ótima ideia 🍌	<a href="https://twitter.com/AlexandreLeony/status/1371156201451311104">https://twitter.com/AlexandreLeony/status/1371156201451311104</a>
2021-03-14 14:47:39	@j_ramacciotti @alaramaximo @Boghosian @AprendaComTC E fecha depois viu	<a href="https://twitter.com/HugoQueiroz_/status/1371156081662164994">https://twitter.com/HugoQueiroz_/status/1371156081662164994</a>
2021-03-14 14:47:31	@j_ramacciotti @alaramaximo @Boghosian @AprendaComTC 🍌🍌🍌🍌🍌🍌🍌🍌🍌	<a href="https://twitter.com/HugoQueiroz_/status/1371156049076584453">https://twitter.com/HugoQueiroz_/status/1371156049076584453</a>
2021-03-14 14:43:06	@alaramaximo @Boghosian @AprendaComTC Kkkkkkkk. Tive que dar um Google	<a href="https://twitter.com/j_ramacciotti/status/1371154936000614408">https://twitter.com/j_ramacciotti/status/1371154936000614408</a>
« Prev 1 2 3 4 Next »		

Fonte: Autor

Esse painel é interessante para que o usuário possa ver os tweets sem ter que entrar no Twitter - ele pode acompanhá-los pela interface do Splunk e ter uma visão imediata dos sentimentos mais recentes. Caso o usuário queira investigar com mais profundidade um tweet, o contexto dele ou seu autor, o link para o tweet original é exibido em uma coluna separada.

A imagem abaixo mostra os um painel que exhibe os tweets mais recentes que foram considerados como “positivos”:

**Figura 36 – Paineis com tweets positivos mais recentes**

Tweets (possivelmente) positivos mais recentes

**Tweets positivos**

_time ↕	tweet ↕	link ↕
2021-03-14 18:30:12	Analisando Guararapes #GUAR3 e vendo o que temos para essa semana no TC Matrix. Enquanto isso, ouvindo aquela rádio na TC STATION do meu TC @AprendaComTC que é show de bola. Daqui a pouco isso vira um SuperApp. Dá pra ficar rapidamente bem informado e recebendo dicas diárias.	<a href="https://twitter.com/xpluis/status/1371212090606780417">https://twitter.com/xpluis/status/1371212090606780417</a>
2021-03-14 10:51:44	Mais uma inovação do @AprendaComTC, ficou muito bom. <a href="https://t.co/QBt5IU6SN">https://t.co/QBt5IU6SN</a>	<a href="https://twitter.com/igorfleitee/status/1371096712819920897">https://twitter.com/igorfleitee/status/1371096712819920897</a>
2021-03-13 15:36:34	Obrigado pela audiência!! 65 ativos abordados na live no @AprendaComTC Bom final de semana! 🦉👍👍 <a href="https://t.co/Or19hLI3Sj">https://t.co/Or19hLI3Sj</a>	<a href="https://twitter.com/sergiosanita/status/1370806005777448965">https://twitter.com/sergiosanita/status/1370806005777448965</a>
2021-03-13 11:35:31	Presidente @FHC na estreia da TC Rádio. Tá bom ou quer mais? @MoverTC @AprendaComTC	<a href="https://twitter.com/gboldrini/status/1370745340392013825">https://twitter.com/gboldrini/status/1370745340392013825</a>
2021-03-13 08:18:19	Imperdível! Reunião Ministerial já foi fantástica, imagina essa sequência que vem no TV Rádio! @AprendaComTC é muito barato!!!! @MoverTC é melhor que Bloomberg!	<a href="https://twitter.com/DrJosevalladao/status/1370695716943843329">https://twitter.com/DrJosevalladao/status/1370695716943843329</a>
2021-03-12 16:37:55	@OgroWallSt @cafeacomferri @AprendaComTC Que legal, gosto muito do trabalho dos dois, parabéns pela reconciliação.	<a href="https://twitter.com/profjoel2/status/1370459054947061763">https://twitter.com/profjoel2/status/1370459054947061763</a>
2021-03-12 11:42:56	@cafeacomferri @AprendaComTC Show de bola. Cheguei no início do slack. N cheguei a pegar o whatsapp, mas mt bom.	<a href="https://twitter.com/Victinhu/status/1370384819448844291">https://twitter.com/Victinhu/status/1370384819448844291</a>
2021-03-12 11:29:21	@AprendaComTC Música para meus ouvidos: selic + 0,75%, dólar 4,80; bolsa 160 mil pts, vacinas, escolas e shoppings abertos, exportações de produtos com + valor agregado, economia crescendo e principalmente BSB trabalhando 9 to 5 de segunda a sexta feira. Um bom jazz tocando.	<a href="https://twitter.com/mariorayel/status/1370381400843677697">https://twitter.com/mariorayel/status/1370381400843677697</a>
2021-03-12 11:21:18	@wanialima @AprendaComTC @BTGPDigital Sua carteira tá melhor que a minha, mãe!	<a href="https://twitter.com/isabelablima/status/1370379375976284164">https://twitter.com/isabelablima/status/1370379375976284164</a>
2021-03-12 11:16:05	Feliz demais com o torneio "Batalha de Titãs" do @AprendaComTC @BTGPDigital aprendendo demais, logo chego em primeiro lugar! 🏆👍 <a href="https://t.co/6Qz96Rql2m">https://t.co/6Qz96Rql2m</a>	<a href="https://twitter.com/wanialima/status/1370378064434192386">https://twitter.com/wanialima/status/1370378064434192386</a>

« Prev 1 2 Next »

Fonte: Autor

Como a imagem demonstra, todos os tweets apresentados são elogios ou comentários positivos sobre a marca. Isso indica que a categorização funciona satisfatoriamente bem. No entanto, a categorização é suscetível a captar elogios sobre coisas alheias e acreditar que tratam-se de elogios ao alvo monitorado.

Um tweet “positivo” é um tweet que contenha qualquer uma das palavras grifadas na imagem abaixo, que mostra o código fonte deste painel:

**Figura 37 – Código fonte do painel de tweets positivos mais recentes**

```
index=main sourcetype="twitter"
| eval datefull = datestamp + " " + mindex(timestamp,1)
| eval _time = strftime(datefull, "%Y-%m-%d %H:%M:%S")
| sort -_time
| where _time > relative_time(now(), "-7d@d")
| search parabens OR legal OR otimo OR ótimo OR bom OR fenomenal OR incrível OR parabéns OR show OR incrível OR feliz OR melhor
| table _time, tweet, link
```

Fonte: Autor

A imagem abaixo mostra um painel que exibe os tweets mais recentes que foram considerados como “negativos”, censurada para remover as palavras ofensivas:

**Figura 38** – Pannel com tweets negativos mais recentes

Tweets (possivelmente) negativos mais recentes

tweets negativos

_time ↕	tweet ↕	link ↕
2021-03-13 07:50:50	@OgroWallSt @cafeacomferri @AprendaComTC "Quem falar [REDACTED] pra mim vai tomar block..." [REDACTED] vira homem [REDACTED]! Cara velho parecendo geração nutella! Bloqueia [REDACTED]!!!!	<a href="https://twitter.com/Anderso24261047/status/1370688801014747140">https://twitter.com/Anderso24261047/status/1370688801014747140</a>
2021-03-12 22:48:27	@wanialima @AprendaComTC @BTGPDigital Oi waniinha cual estrategia eh essa? Eh aquela "segue o ferri ele embolsa o lucro e a gente morre com as [REDACTED] na carter?" Ja participei desa hoje meu filho nao tem vidio gueime porque eu perdi muito com conga, foge deles eles sao do mal nao ajudam quando vc quebra	<a href="https://twitter.com/Crypto_Menas/status/1370552302654926853">https://twitter.com/Crypto_Menas/status/1370552302654926853</a>
2021-03-12 15:21:42	@cafeacomferri @AprendaComTC Diferenças e tretas à parte @cafeacomferri , você me inspira bastante nessa área! Sério! Obs.: Quem falar [REDACTED] pra mim, vai tomar block.	<a href="https://twitter.com/OgroWallSt/status/1370439874373881856">https://twitter.com/OgroWallSt/status/1370439874373881856</a>
2021-03-11 23:14:54	@vcamargooo @Tatobarros @albuquerque_af @cafeacomferri @omarajame @AprendaComTC [REDACTED] HAHAAHHA	<a href="https://twitter.com/Pedrfw/status/1370196570717749250">https://twitter.com/Pedrfw/status/1370196570717749250</a>
2021-03-11 21:09:09	@alinezeitadiaz @Tatobarros @AprendaComTC @cafeacomferri @mayvalentim2 @israel_massa FHC no TC. Que [REDACTED]!!!	<a href="https://twitter.com/LzaroCoutinho5/status/1370164927588143104">https://twitter.com/LzaroCoutinho5/status/1370164927588143104</a>
2021-03-11 19:51:32	@vcamargooo @Tatobarros @albuquerque_af @cafeacomferri @omarajame @AprendaComTC Ahahahahah [REDACTED] bagunça	<a href="https://twitter.com/jfcbarrros/status/1370145395389648900">https://twitter.com/jfcbarrros/status/1370145395389648900</a>
2021-03-11 19:47:20	@Tatobarros @vcamargooo @albuquerque_af @cafeacomferri @omarajame @AprendaComTC HAHAAHHA [REDACTED] HAHAAHHA	<a href="https://twitter.com/igor_oliveira/status/1370144337372315648">https://twitter.com/igor_oliveira/status/1370144337372315648</a>
2021-03-11 11:19:09	@AprendaComTC @albuquerque_af @cafeacomferri @Tatobarros @omarajame Só para constar.. eu vou estar colado ouvindo Tô na [REDACTED] mesmo... literalmente minha conta não tem mais downside 😞😞😞 Se cair mais vai para o inferno Então... me ajudem a sair dessa e façam uma reunião com boas dicas de mercado	<a href="https://twitter.com/tonabruxa/status/1370016448794476547">https://twitter.com/tonabruxa/status/1370016448794476547</a>

Fonte: Autor

Tendo em vista a quantidade de palavras censuradas, a categorização parece funcionar bem, mas tende a categorizar comentários autodepreciativos, humorísticos ou que contenham palavras ofensivas como “negativos” contra a marca. O tweet na parte inferior da imagem é um ótimo exemplo disso: o usuário faz um comentário negativo sobre si mesmo usando uma palavra de baixo calão, e a categorização entende isso como uma menção negativa contra a marca.

Um tweet negativo é um tweet que contenha qualquer uma das palavras grifadas na imagem abaixo, que mostra o código fonte desse painel:

**Figura 39** – Código fonte do painel de tweets negativos mais recentes

```
index=main sourcetype="twitter"
| eval datefull = timestamp + " " + mvindex(timestamp,1)
| eval _time = strptime(datefull, "%Y-%m-%d %H:%M:%S")
| where _time > relative_time(now(), "-30d@d")
| sort _time
| search merda OR bosta OR pessimo OR péssimo OR horrivel OR horroroso OR lixo OR escroto OR arrombado OR arrombados
OR puta OR decepcionado OR decepcionante OR cancelado OR cancelar OR horrivel OR lento OR injusto OR criminoso
| table _time, tweet, link
```

Fonte: Autor

É claro que as duas listas de palavras são facilmente modificáveis e podem ser melhor configuradas para marcas específicas e o vocabulário comum dos usuários que interagem com cada uma.



Um exemplo interessante foi capturado onde dois clientes aparentemente muito insatisfeitos reclamaram dos serviços da empresa:

**Figura 40** – Exemplos de tweets negativos identificados

Tweets (possivelmente) negativos mais recentes		
tweets negativos		
_time ↕	tweet ↕	link ↕
2021-02-23 11:04:24	@cafecomferri @IsraelSalmen @AprendaComTC Acho que vou cancelar o TC. Tudo acontece no Clubhouse agora...	<a href="https://twitter.com/xopenh/status/1364214531287904256">https://twitter.com/xopenh/status/1364214531287904256</a>
2021-02-20 07:26:32	@isabelablima @albuquerque_af @AprendaComTC Tc é um lixo. Plataforma fraca, cobram uma fortuna e não oferecem nada. Um bando de lunáticos colocando operações loucas e sem noção.	<a href="https://twitter.com/AnaKarinaArq/status/1363072540592193538">https://twitter.com/AnaKarinaArq/status/1363072540592193538</a>

Fonte: Autor

Esse é um caso em que, se o usuário configurar um alerta para tweets negativos, poderia receber uma notificação sobre esses tweets e entrar em contato com os usuários para tentar reparar a relação entre a empresa e o cliente. O UBM não vem com alertas pré-configurados, é necessário que o usuário os crie.

Se o usuário seguir um tutorial introdutório de SPL ele pode começar a modificar as queries fornecidas para aprofundar suas buscas, extrair outras formas de resultados ou inteligências, ou criar mais painéis para o dashboard. Por exemplo, podemos facilmente usar a query de tweets negativos ou positivos para buscar por outras palavras. Abaixo exibimos uma query básica criada em menos de 5 minutos para buscar todos os tweets com palavras relacionadas a bugs e falhas:

**Figura 41** – Busca modificada para pesquisar tweets mencionando problemas técnicos

New Search

Save As Close

```

1 index=main sourcetype="twitter"
2 | eval datefull = timestamp + " " + mvindex(timestamp,1)
3 | eval _time = strftime(datefull, "%Y-%m-%d %H:%M:%S")
4 | sort -_time
5 | search bug OR crash OR bugado OR crashando OR bugou OR "não funciona" OR "nao funciona"
6 | table _time, tweet, link

```

All time

Q

✓ 6 events (before 3/14/21 8:48:58.000 PM) No Event Sampling

Job II ↗ ↘ ⬇ ⬆ ⬇ ⬆ Fast Mode

Events Patterns Statistics (6) Visualization

20 Per Page Format Preview

_time	tweet	link
2021-02-18 21:34:15	@luis_sayeg @cafeconferri @albuquerque_af @AprendaComTC @sergiosanita Trocou o fone e não resolveu, é um bug no instagram pelo que parece.	<a href="https://twitter.com/israel_massa/status/1362561098004717568">https://twitter.com/israel_massa/status/1362561098004717568</a>
2020-12-07 22:10:47	@tradersclubbr @femisapien_z Puts só ouvi promessas que não foram entregues de vocês!!!! Prefiro voltar pro meu excel do que o app de vcs que so da bug. Alias, pq pararam de falar de conga? Levei vara..	<a href="https://twitter.com/CAPITAAGR/status/1336115978141372416">https://twitter.com/CAPITAAGR/status/1336115978141372416</a>
2020-10-23 12:51:37	@tradersclubbr bugou geral e o site caiu... aiaiaiai	<a href="https://twitter.com/MisterChefBR/status/1319667802723274752">https://twitter.com/MisterChefBR/status/1319667802723274752</a>
2020-08-23 17:32:23	@tradersclubbr Ta com bug nos relatórios do BTG, quando clico ele desloga do app do tc! Vou abrir um chamado.. se tiver uma vaga de QA também aceito enviar meu CV hehehe	<a href="https://twitter.com/AplicarNaBolsa/status/1297632801370628109">https://twitter.com/AplicarNaBolsa/status/1297632801370628109</a>
2020-08-06 14:42:02	@tradersclubbr @cafeconferri @albuquerque_af @israel_massa @claupimenta2 @isabelablima @diiandrasantos @UhlighLucas Já fiz 'n' videos sobre esse BUG, mandei aqui no twitter 'n' vezes, abri chamado, saiu atualização, o BUG continua, mais e mais gente reclamando... 🤖🤖 <a href="https://t.co/WT2Yr9jHmJ">https://t.co/WT2Yr9jHmJ</a>	<a href="https://twitter.com/AlbieroFelipe/status/1291429339880398849">https://twitter.com/AlbieroFelipe/status/1291429339880398849</a>
2020-07-26 20:22:22	@tradersclubbr Por favor corrijam o BUG do Tradezoom com IVB11, não funciona.	<a href="https://twitter.com/circuit23696156/status/1287528720060821505">https://twitter.com/circuit23696156/status/1287528720060821505</a>

Fonte: Autor

Isso também mostra que o UBM pode ser utilizado como uma ferramenta de busca dentre todas as menções disponíveis.

## **5 – CONCLUSÕES E TRABALHOS FUTUROS**

Conseguimos criar uma solução gratuita (apelada de UBM) para usos pequenos, em arquitetura distribuída, open-source, modular, escalável e de fácil utilização que permite a monitoração de um sujeito, organização, instituição ou qualquer outro alvo e apresenta ao usuário uma visão automaticamente mantida da imagem pública dos alvos monitorados, junto com uma gama de informações úteis.

A solução pode ser instalada facilmente e por qualquer indivíduo, não exigindo um conhecimento técnico avançado. Ela também pode ser hospedada em diferentes formatos, tanto para o Splunk (seu banco de dados e motor de buscas) quanto para os coletores de dados de mídias sociais.

Devido à arquitetura distribuída, ela é facilmente extensível - usuários podem criar coletores, dashboards e queries e facilmente compartilhá-los com outros usuários. Um usuário pode usar apenas os coletores que quiser, simplesmente não instalando os que não deseja.

Acreditamos que esse trabalho atingiu os objetivos listados na seção 1.3.

Para trabalhos futuros, acreditamos que o desenvolvimento de coletores, queries e dashboards para outras redes sociais como Facebook, Instagram e Reddit seria interessante, pois poderiam aumentar a quantidade de informações brutas disponíveis para tratamento e consequentemente a quantidade de inteligência e visibilidade gerada pela aplicação.

Também poderiam ser criadas mais queries interessantes para os dados do próprio twitter: informações como seguidores mais ativos, críticas e elogios com mais likes, comentários de pessoas “famosas” no twitter (com mais de um certo número de seguidores) e afins podem ser indicadores interessantes.

Outra frente interessante para trabalhos futuros é a integração de sistemas PLN (Processamento de Linguagem Natural) para uma categorização mais efetiva entre comentários positivos ou negativos aos termos monitorados.

## **6 - REFERÊNCIAS BIBLIOGRÁFICAS**

- [1] PANIAGUA, Jordi; SAPENA, Juan. Business performance and social media: Love or hate?. **Business horizons**, v. 57, n. 6, p. 719-728, 2014.
- [2] HERBIG, Paul; MILEWICZ, John. The relationship of reputation and credibility to brand success. **Journal of consumer marketing**, 1993.
- [3] ATAMAN, Berk; ÜLENGİN, Burç. A note on the effect of brand image on sales. **Journal of Product & Brand Management**, 2003.
- [4] DISTASO, Marcia W.; MCCORKINDALE, Tina; WRIGHT, Donald K. How public relations executives perceive and measure the impact of social media in their organizations. **Public relations review**, v. 37, n. 3, p. 325-328, 2011.
- [5] <https://mention.com/en/pricing/>. Acesso em 18 mar. 2021
- [6] <https://www.agorapulse.com/features/social-media-monitoring/>. Acesso em 18 mar. 2021
- [7] <https://www.google.com/alerts>. Acesso em 19 mar. 2021
- [8] <https://www.splunk.com/>. Acesso em 19 mar. 2021
- [9] <https://dev.splunk.com/enterprise/docs/devtools/python/sdk-python/>. Acesso em 19 mar. 2021
- [10] <https://github.com/lacioffi/UniversalBrandMonitor>. Acesso em 20 mar. 2021
- [11] <https://twitter.com/aprendacomtc>. Acesso em 20 mar. 2021