

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
INSTITUTO TERCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

LEANDRO DA SILVA GONÇALVES

DISSERTAÇÃO DE MESTRADO

UM FRAMEWORK CONCEITUAL PARA APOIAR O DESENVOLVIMENTO DE
SOLUÇÕES QUE UTILIZEM AS INFORMAÇÕES DE MÍDIAS SOCIAIS PARA
DETECÇÃO DE RUMORES EM SAÚDE PÚBLICA

RIO DE JANEIRO
2022

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

LEANDRO DA SILVA GONÇALVES

UM FRAMEWORK CONCEITUAL PARA APOIAR O DESENVOLVIMENTO DE
SOLUÇÕES QUE UTILIZEM AS INFORMAÇÕES DE MÍDIAS SOCIAIS PARA
DETECÇÃO DE RUMORES EM SAÚDE PÚBLICA

Dissertação apresentada ao Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais da Universidade Federal do Rio de Janeiro – PPGI/UFRJ, como parte dos requisitos necessários à obtenção o Título de Mestre em Informática (Metodologias e Técnicas de Computação – MT).

Orientador: Prof. Paulo Victor Rodrigues de Carvalho

Coorientador: Prof. Danilo Pestana de Freitas

Rio de Janeiro
2022

CIP - Catalogação na Publicação

G635f Gonçalves, Leandro da Silva
Um framework conceitual para apoiar o desenvolvimento de soluções que utilizem as informações de mídias sociais para detecção de rumores em saúde pública / Leandro da Silva Gonçalves. -- Rio de Janeiro, 2022.
96 f.

Orientador: Paulo Victor Rodrigues de Carvalho.
Coorientador: Danilo Pestana de Freitas.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Programa de Pós-Graduação em informática, 2022.

1. Mídias sociais. 2. CIEVS. 3. Vigilância em saúde. 4. Rumor. 5. Twitter. I. Carvalho, Paulo Victor Rodrigues de, orient. II. Freitas, Danilo Pestana de, coorient. III. Título.

LEANDRO DA SILVA GONÇALVES

UM FRAMEWORK CONCEITUAL PARA APOIAR O DESENVOLVIMENTO DE
SOLUÇÕES QUE UTILIZEM AS INFORMAÇÕES DE MÍDIAS SOCIAIS PARA
DETECÇÃO DE RUMORES EM SAÚDE PÚBLICA

Dissertação de Mestrado submetida ao Programa de Pós-graduação em Informática do Instituto de Matemática e do Instituto Tércio Pacitti da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Mestre em Informática.

Aprovada em: 15 de maio de 2022 por:



Prof. Paulo Victor Rodrigues de Carvalho, D.Sc.,
PPGI/UFRJ (Presidente)

“participação por vídeo conferência”

Prof. Danilo Pestana de Freitas, D.Sc, FACC/UFRJ

”participação por vídeo conferência”

Prof.^a. Mônica Ferreira da Silva, D.Sc., PPGI/UFRJ

“participação por vídeo conferência”

Prof. José Orlando Gomes, D.Sc., PPGI/UFRJ

“participação por vídeo conferência”

Prof. Hugo Cesar Bellas, D.Sc., FIOCRUZ

“A persistência é o menor caminho do êxito”

(Charles Chaplin)

AGRADECIMENTOS

Um trabalho de mestrado representa uma longa trajetória cheia de desafios, tristezas, incertezas, alegrias e percalços. Mas apesar do processo solitário a que qualquer pesquisador está destinado, reúne contribuições de várias pessoas, indispensáveis para encontrar o melhor rumo a cada passo do caminho percorrido. Percorrer essa trilha só foi possível com apoio, energia e forças de várias pessoas, a quem dedico especialmente este projeto.

Especialmente a minha esposa e grande companheira, Patrícia, agradeço o apoio, motivação incondicional e encorajamentos em momentos cruciais que ajudaram a tornar este trabalho uma válida e agradável experiência de aprendizagem.

A minha família. Meus pais Dair e Gerceleste e minha irmã Mônica pelos conselhos preciosos, companheirismo e enorme compreensão que contribuíram para chegar ao final deste percurso. Minha tia Sônia, minha sogra Regina e cunhada Thaís, grandes incentivadoras e entusiastas deste projeto. Aos meus pets Maya e Meghan, que me acompanharam em diversos momentos desta caminhada.

Aos amigos especiais como: Karen, Sheila e Marco, que acompanharam de perto e souberam ser pacientes e incentivadores desta jornada.

A todos os professores do PPGI, muito obrigado por cada aula. O aprendizado adquirido em cada uma delas foi crucial para construir este projeto.

Aos meus orientadores Paulo Victor Rodrigues de Carvalho e Danilo Pestana de Freitas, que sempre acreditaram em mim, muito obrigado por cada reunião, cada revisão, cada dica. A orientação exemplar pautada por um elevado e rigoroso nível científico, uma visão crítica e oportuna, um empenho e exigência na medida certa, os quais contribuíram para enriquecer, com grande dedicação, cada uma das etapas deste trabalho.

Por fim, a todos aqueles que contribuíram, direta ou indiretamente, para realização desta dissertação, o meu sincero agradecimento.

RESUMO

Nas últimas décadas o comportamento epidemiológico dos países tem apresentado mudanças, tanto nos padrões de ocorrências de doenças infecciosas quanto nos seus respectivos padrões de transmissão, dificultando a mitigação da ocorrência de possíveis crises em saúde pública. Por isso, em 2005, a OMS revisou o Regulamento Sanitário Internacional para realização de diversas atualizações em seu conteúdo, entre elas a inclusão de novas abordagens no que se refere à vigilância de doenças e riscos. Motivados por esta revisão, diversos países realizaram ajustes em suas estratégias de vigilância em saúde, incluindo o Brasil. Entre outras ações, este país, membro da OMS (Organização Mundial de Saúde), implementou os Centros de Informações Estratégicas em Vigilância em Saúde Pública (CIEVS) para atuarem na detecção, investigação e coordenação da resposta a uma possível ameaça à saúde pública, que seriam responsáveis por rastrear informações relevantes sobre saúde, tanto de fontes oficiais, quanto não oficiais.

Ao mesmo tempo, as Mídias Sociais vêm se mostrando uma boa fonte de informações não oficiais relacionadas a saúde, pois, cada vez, mais seus usuários estão publicando ideias, discussões e informações sobre sua própria condição de saúde ou da comunidade em que vive. Tais características chamam a atenção das equipes de vigilância em saúde para utilizar estas informações, porém há a dificuldade de extração, tratamento e disponibilização oportuna de informações relevantes para suas atividades. Para agilizar este processo, os recursos de Tecnologia da Informação são cada vez mais requeridos, pois são capazes de implementar soluções sistêmicas e automatizadas capazes de tratar este grande volume de informações das Mídias Sociais.

O objetivo desta pesquisa foi a proposição de uma estrutura de análise das informações de Mídias Sociais que facilite o seu uso pelas equipes responsáveis pela detecção de rumores em saúde pública. Para isso, foi realizada uma pesquisa bibliográfica para mapear o estado da arte sobre o funcionamento de vigilância em saúde no Brasil, o conhecimento obtido foi posteriormente aprofundado com a realização de uma entrevista com especialistas e profissionais em saúde que explicaram sobre as estratégias abordadas e como ocorre o uso das informações de Mídias Sociais no processo de detecção de rumores em saúde pública.

Após as elucidações obtidas, esta pesquisa elaborou um Framework conceitual para representar a análise das informações de Mídias Sociais e para direcionar o desenvolvimento de sistemas que apoiem as equipes no processo de detecção de rumores. Após sua construção, o Framework foi apresentado aos especialistas e profissionais em saúde para coleta de suas percepções e identificação de oportunidades de melhoria do artefato. Finalizado o Framework, houve a elaboração de duas instâncias para simular investigações distintas sobre a existência de rumores das síndromes diarreica e respiratória com uso de informações do Twitter. Os resultados obtidos foram apresentados e avaliados pelos especialistas e profissionais envolvidos que observaram uma redução no esforço em investigações similares. Ao investigar existência de rumores sobre síndrome diarreica e respiratória, os resultados exibiram 2,32 % e 6,17%, respectivamente, de mensagens com relevância segundo os parâmetros aplicados.

Os resultados apresentados neste trabalho, em todas as fases, foram suficientes para mostrar que através do estabelecimento de uma estrutura, com fases distintas e parametrizadas, que permita o emprego de lógica em seus parâmetros de análise e refinamentos, pode reduzir o esforço realizado pelas equipes ao utilizar as informações de Mídias Sociais para identificação de rumores em saúde pública.

Palavras-chave: Mídias Sociais; CIEVS; Vigilância em Saúde; rumor; Twitter.

ABSTRACT

In recent decades, the epidemiological behavior of countries has shown changes, both in the patterns of occurrence of infectious diseases and in their respective patterns of transmission, making it difficult to mitigate the occurrence of possible crises in public health. For this reason, in 2005, the WHO revised the International Health Regulations to carry out several updates to their content, including the inclusion of new approaches in terms of disease and risk surveillance. Motivated by this review, several countries made adjustments to their health surveillance strategies, including Brazil. Among other actions, this country, a member of the WHO (World Health Organization), implemented the Strategic Information Centers on Public Health Surveillance (CIEVS) to act in the detection, investigation and coordination of the response to a possible threat to public health, which would be responsible for tracking relevant health information from both official and unofficial sources.

At the same time, Social Media has proved to be a good source of unofficial health-related information, as more and more of its users are publishing ideas, discussions and information about their own health condition or that of the community in which they live. Such characteristics draw the attention of health surveillance teams to use this information, but there is difficulty in extracting, processing and timely availability of information relevant to their activities. To speed up this process, Information Technology resources are increasingly required, as they are capable of implementing systemic and automated solutions capable of handling this large volume of information from Social Media.

The objective of this research was to propose a structure for the analysis of Social Media information that facilitates its use by the teams responsible for detecting rumors in public health. For this, a bibliographic research was carried out to map the state of the art on the functioning of health surveillance in Brazil, the knowledge obtained was later deepened with an interview with specialists and health professionals who explained about the strategies addressed and how Social Media information is used in the process of detecting rumors in public health.

After the elucidations obtained, this research developed a conceptual framework to represent the analysis of Social Media information and to guide the development of systems that support teams in the process of detecting rumors. After its construction, the Framework was presented to specialists and health professionals to collect their perceptions and identify opportunities to improve the artifact. Once the Framework was completed, two instances were created to simulate different investigations into the existence of rumors of diarrheal and respiratory syndromes using information from Twitter. The results obtained were presented and evaluated by the specialists and professionals involved who observed a reduction in the effort in similar investigations. When investigating the existence of rumors about diarrheal and respiratory syndrome, the results showed 2.32% and 6.17%, respectively, of messages with relevance according to the applied parameters.

The results presented in this work, in all phases, were enough to show that through the establishment of a structure, with distinct and parameterized phases, which allows the use of logic in its parameters of analysis and refinements, the effort made by the teams can be reduced. when using Social Media information to identify public health rumors.

Keywords: Social Media; CIEVS; Health Surveillance; rumor; Twitter.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estrutura CIEVS no Brasil.....	21
Figura 3 - Visão Geral de Todas as Funções de Vigilância e Resposta à Saúde Pública.....	23
Figura 4 - Vigilância Baseada em Indicadores, Baseado em Eventos e Inteligência Epidêmica	24
Figura 5 - Método da Pesquisa	39
Figura 6 - Processo de Desenvolvimento de um Framework.....	42
Figura 7 - Estrutura do Framework	45
Figura 8 - Fluxograma para elaboração do Plano de Captura	47
Figura 9 - Fluxograma para elaboração do Plano de Análise.....	49
Figura 10 - Descarte de Mensagens por Período.....	52
Figura 11 - Descarte por Posicionamento Geográfico.....	53
Figura 12 - Fluxo de Reanálise das Mensagens	54
Figura 13 - Exemplo de Relatório	55
Figura 14 - Classificações das Mensagens	56
Figura 15 - AP's do Rio de Janeiro	67
Figura 16 - Plano de Captura - Área de Abrangência.....	68

LISTA DE GRÁFICOS

Gráfico 2 - Frequência de Busca por Informações nas Mídias Sociais	26
Gráfico 3 - Distribuição dos casos notificados em unidades sentinelas por região e ano, Brasil 2009 a 2018	63
Gráfico 4 - Casos de SRAG por faixa etária	64
Gráfico 6 – Síndrome Diarreica - Mensagens Capturadas por Dia	71
Gráfico 6 – Síndrome Respiratória - Mensagens Capturadas por Dia	76
Gráfico 8 - Relação Palavras-Chave e Mensagens Capturadas nas Instâncias.....	80
Gráfico 8 - Mensagens das Instâncias	81

LISTA DE TABELAS

Tabela 1: Questionário da entrevista	40
Tabela 2 - Exemplos de Lógicas Sindrômicas	50
Tabela 3 - Exemplo de Tweets Descartados.....	51
Tabela 4 - Compilação da Avaliação conceitual do Framework.....	60
Tabela 5 - Organização das palavras de Captura.....	66
Tabela 6 – Síndrome Diarreica -Plano de Captura-Palavras-Chave.....	70
Tabela 7 – Síndrome Diarreica -Lógica Sindrômica	72
Tabela 8 - Parâmetros de Descarte	73
Tabela 9 - Classificações - Resumo Geral.....	74
Tabela 10 – Síndrome Respiratória-Plano de Captura	75
Tabela 11 - SR - Lógica Sindrômica - Palavras-Chave.....	77
Tabela 12 - Parâmetros de Descarte	78
Tabela 13 - Classificações - Resumo Geral.....	79

LISTA DE SIGLAS

API	Application Programming Interface
ESP	Evento em Saúde Pública
ESPII	Evento em Saúde Pública de Importância Internacional
ESPIN	Evento em Saúde Pública de Importância Nacional
GOARN	Rede Global de Alerta e Resposta a Surtos
IE	Inteligência Epidêmica
MS	Ministério da Saúde
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-americana da Saúde
RSI	Regulamento Sanitário Internacional
SARS	Síndrome Respiratória Aguda Grave
SVS	Secretaria de Vigilância em Saúde
VBE	Vigilância Baseada em Eventos
VBI	Vigilância Baseada em Indicadores

SUMÁRIO

1	INTRODUÇÃO	13
1.1	MOTIVAÇÃO DA PESQUISA	13
1.2	PROBLEMA DE PESQUISA	15
1.3	OBJETIVO DA PESQUISA	16
1.4	RELEVÂNCIA DA PESQUISA	16
1.5	ORGANIZAÇÃO DO TRABALHO	17
2	REVISÃO DE LITERATURA	19
2.1	RESPOSTAS ÀS EMERGÊNCIAS EM SAÚDE PÚBLICA NO BRASIL	19
2.2	DA VIGILÂNCIA EM SAÚDE ATÉ A INTELIGÊNCIA EPIDÊMICA	22
2.3	MÍDIAS SOCIAIS NA VIGILÂNCIA EM SAÚDE	26
3	MÉTODO DE PESQUISA	37
4	ENTREVISTA NÃO ESTRUTURADA	40
5	FRAMEWORK PARA APOIO À DETECÇÃO DE RUMORES	42
5.1	FRAMEWORK: VISÃO GERAL	44
5.2	FUNÇÕES DO FRAMEWORK	46
5.2.1	Planejamento da captura de rumores em saúde	46
5.2.2	Captura das mensagens	48
5.2.3	Tratamento das mensagens	48
5.2.3.1	Análise sindrômica	49
5.2.3.2	Descarte de mensagens	50
5.2.3.3	Reanálise das mensagens	53
5.2.4	Alertas e relatórios	54
5.2.5	Classificação das mensagens	55
5.2.6	Repositórios do framework	57
6	AVALIAÇÃO CONCEITUAL DO FRAMEWORK	59
6.1	INSTANCIAMENTO DO FRAMEWORK	62
6.1.1	Estrutura das instâncias	64
6.1.1.1	Plano de captura	65
6.1.1.2	Plano de análise	69
6.1.2	Identificação de rumor sobre síndrome diarreica	70
6.1.3	Identificação de rumor sobre síndrome respiratória	75
6.2	ANÁLISE DOS RESULTADOS	80

6.2.1	Captura das mensagens	80
6.2.2	Análise das mensagens	81
6.2.3	Emissões de alertas	81
7	DISCUSSÃO	82
8	CONCLUSÃO.....	84
8.1	CONTRIBUIÇÕES	84
8.2	LIMITAÇÕES	85
8.3	TRABALHOS FUTUROS	86
	REFERÊNCIAS	87

1 INTRODUÇÃO

Neste capítulo são apresentadas a motivação, o problema, o objetivo, a relevância e a organização do trabalho.

1.1 MOTIVAÇÃO DA PESQUISA

Nos últimos anos, a identificação de mudanças nos padrões de ocorrências de doenças infecciosas e na dinâmica de transmissão de seus agentes infecciosos, tem gerado reflexões sobre os consideráveis desafios que os sistemas de vigilância em saúde pública têm que enfrentar. Fatores como: a universalização das transações comerciais e repercussões econômicas; ampliação de mercados e acordos de livre comércio; livre fluxo de pessoas; o aumento da densidade populacional urbana; mudanças no comportamento social e até mesmo alterações no meio ambiente, influenciaram na ocorrência de diversas crises na saúde pública (CARMO, 2020; FORTES; RIBEIRO, 2014; GRISOTTI, 2010; MATOS, 2011).

Representadas não só pelos registros de epidemias e pandemias ocasionadas por novas doenças com perfil epidemiológico ainda desconhecido ou pelo retorno de doenças já conhecidas (LIMA, 2002) essas crises em saúde pública, compreendem também o surgimento de doenças modificadas que aumentam o seu contágio ou elevam a sua resistência a medicamentos dos agentes já conhecidos (TEIXEIRA; COSTA; CARMO; *et al.*, 2018; TEIXEIRA; COSTA; PAIXÃO; *et al.*, 2018). Essas modificações ocasionaram grandes alterações nos perfis epidemiológicos e nas dinâmicas de transmissão já conhecidas para essas doenças, que contribuem para dificultar ainda mais a mitigação de suas ocorrências como a Síndrome Respiratória Aguda Grave (SARS) em 2003, a pandemia de influenza H1N1 em 2009 e o Ebola entre 2013 e 2016 (GUI *et al.*, 2017).

Para o enfrentamento destes desafios, algumas mudanças de estratégias tiveram que ser consideradas, para auxiliar na elevação do desempenho dos sistemas de vigilância em saúde, como por exemplo: a complementação das informações provenientes dos sistemas de vigilância tradicionais com outras informações consideradas não oficiais (rumores) (PAQUET *et al.*, 2006); a adoção de uma nova abordagem concentrada nos eventos em saúde pública, apoiando a atual vigilância através de indicadores e relatórios (BARBOZA *et al.*, 2013); estabelecimento de uma abordagem mais analítica, sobre as informações

disponibilizadas pela internet, utilizando-as como ferramenta de apoio tanto para prestar informações sobre saúde, quanto para capturar informações que possam ser úteis ao processo de vigilância (EYSENBACH, 2006; MAVRAGANI, 2020).

No Brasil, através dos Centros de Informações Estratégicas em Vigilância em Saúde (CIEVS), foram implementadas algumas dessas estratégias. Importantes componentes do Ministério da Saúde, os CIEVS atuam diretamente no monitoramento, investigação e coordenação da resposta a uma possível ameaça à saúde pública. Para realização destas atividades, eles monitoram diversas fontes de informações, sejam elas oficiais ou não oficiais. Através dos sistemas de vigilância em saúde tradicionais, os CIEVS obtêm informações importantes para suas atividades e através de processos como Detecção Digital de Doenças (DDD) e Vigilância Participativa são realizados o monitoramento de rumores, tanto em serviços baseados em internet, quanto diretamente da população.

Por mais que o contato direto com a população seja uma boa fonte de informações sobre saúde, os serviços suportados pela internet são ainda melhores, pois eles representam uma enorme fonte de informações oportunas e diversificadas sobre saúde que são disponibilizadas ou compartilhadas por seus usuários sobre si mesmos ou sobre as comunidades a que pertencem (KASS-HOUT; ALHINNAWI, 2013). Isso aumenta o interesse em monitorar surtos de doença através do uso dessas informações que, geralmente são obtidas por meio da aplicação de técnicas para análises de informações e notícias (ALAMOUDI *et al.*, 2021; TEIXEIRA; COSTA; CARMO; *et al.*, 2018; WANG *et al.*, 2021; ZHANG; ZHANG; LI, 2015) e mecanismos de pesquisa e padrões em navegação (ROVETTA; BHAGAVATHULA, 2020).

Definidas por BOYD e ELLISON (2007) como serviços baseados na web que provêm um meio social onde seus usuários são incentivados a se comunicarem através do compartilhamento e criação de novas informações, quase de forma instantânea, as Mídias Sociais se destacam pelo seu potencial no fornecimento de informações não oficiais (rumores) capazes de auxiliar a contextualização sobre os problemas de saúde de uma população local (ABDULLAH; WU, 2011; KLEIN; GUIDI NETO; TEZZA, 2017). Porém, a utilização dessas informações não é uma tarefa trivial, pois o grande volume de informações, a falta de padrões e estrutura da informação e a demora na compreensão da informação útil extraída, podem prejudicar a potencial contribuição para a realização de vigilância em saúde pública (HU; TANG; LIU, 2011).

1.2 PROBLEMA DE PESQUISA

Estudos mostram que medidas preventivas podem ser tomadas para conter possíveis epidemias logo em seu início, através da realização de uma detecção precoce de qualquer evento em saúde pública (FERGUSON *et al.*, 2005; LONGINI, 2005). Mesmo com o subsídio das informações advindas dos atuais sistemas de vigilância sanitária, ainda há precariedade de informações úteis em momento oportuno devido aos atuais processos utilizados. Essas ausências podem representar lacunas na compreensão contextual em um cenário de emergência em saúde pública (XAVIER *et al.*, 2020). Com isso, entende-se que informações disponibilizadas diretamente por cidadãos de uma região podem ajudar em uma melhor contextualização do cenário.

Devido a sua popularidade, as Mídias Sociais são utilizadas para rápidas postagens por seus usuários sobre diversos assuntos, incluindo assuntos de saúde. Este volume de informações sobre saúde encontrada nas Mídias Sociais, representa o seu grande potencial como fonte de rumores úteis ao processo de vigilância em saúde. Estas informações são capazes de trazer novidades ou mesmo orientar as equipes de vigilância a um melhor entendimento de um cenário sob investigação, auxiliar o acompanhamento sobre um comportamento epidemiológico e a detectar precocemente uma possível emergência.

Porém, pela falta do estabelecimento de uma estrutura para o grande volume de informações disponíveis, existe a necessidade do emprego de um grande esforço para identificar informações relevantes e disponibilizá-las oportunamente. O desafio está na identificação de uma estrutura que possa ser adotada para apoiar esse processo, reduzindo o esforço empregado.

Então, a adoção de uma estrutura adequada para as informações de Mídias Sociais, pode auxiliar na obtenção de informações úteis e oportunas que apoiem o processo de detecção de rumores realizado pelas equipes de vigilância em saúde.

Desta forma, esta dissertação propõe responder a seguinte questão:

“Como estruturar as informações das Mídias Sociais para facilitar o seu uso pelas equipes de detecção de rumores em saúde pública?”

1.3 OBJETIVO DA PESQUISA

O objetivo desta pesquisa visa entender o processo de análise e estrutura de informações de Mídias Sociais para possibilitar o desenvolvimento de soluções computacionais, que sejam capazes de aplicar uma estrutura para facilitar o uso destas informações, pelas equipes de investigação de rumores em saúde pública.

Para atingir este objetivo, a pesquisa se desdobra nos seguintes objetivos específicos:

- Conduzir um estudo e entrevistas com especialistas para compreensão do processo de identificação de rumores em saúde pública e como as informações de Mídias Sociais são utilizadas e estruturadas;
- Definir um framework conceitual que auxilie a construção de sistemas de informação capazes de estruturar as informações de Mídias Sociais para reduzir o esforço nas investigações sobre a existência de rumores;
- Conduzir uma avaliação qualitativa do framework junto a profissionais e especialistas da área da saúde;
- Estabelecer uma simulação do uso do framework com dados reais coletados diretamente de uma Mídia Social para realização de uma investigação;
- Conduzir uma avaliação qualitativa das simulações construídas junto a profissionais e especialistas da área da saúde.

1.4 RELEVÂNCIA DA PESQUISA

Com a dinâmica socioeconômica e o conseqüente aumento no fluxo de pessoas e mercadorias pelo mundo, há uma elevação da probabilidade de disseminação de doenças. Como exemplo, temos o novo Coronavírus (COVID-19) que rapidamente registrou-se como a sexta doença na história a ocasionar o status de Emergência na Saúde Pública de Importância Internacional (ESPII), pois trata-se de um evento extraordinário que constitui um risco de saúde pública de proporções internacionais que requer uma resposta coordenada e imediata (ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE, 2020).

Casos como este, exemplificam uma das preocupações antigas das entidades na área da saúde que é a necessidade da realização do acompanhamento do comportamento epidemiológico das doenças (TEIXEIRA; COSTA; CARMO; *et al.*, 2018) para evitar possíveis crises ainda no início. No passado, com intenção de estabelecer padrões mundiais mais adequados e mais capazes para garantir o sucesso do controle de epidemias ou mesmo de pandemias, em 2005, a Organização Mundial da Saúde (OMS) revisou o Regulamento Sanitário Internacional (RSI) (WORLD HEALTH ORGANIZATION, 2005), cujas recomendações são seguidas por todos os seus países membros até os dias de hoje.

Uma das preocupações do novo RSI foi a melhoria da capacidade dos países em detectar precocemente potenciais riscos relacionados à saúde. Então, ao observar que os métodos tradicionais de vigilância podem não fornecer informações oportunas devido a existência de um longo processo de validação, burocracias, alto custo e necessidade de muitos recursos (YANG; HORNEFFER; DILISIO, 2013) o RSI 2005 passou a considerar a utilização de fonte de informações não oficiais, em complemento às informações oficiais (YAN; CHUGHTAI; MACINTYRE, 2017).

No mundo hiper conectado de hoje, boa parte das pessoas com qualquer desconforto realizam autodiagnósticos pela internet, reportando seu estado atual e buscando informações sobre seus sintomas (KASS-HOUT; ALHINNAWI, 2013). Ao fazerem isso, essas mesmas pessoas deixam informações relevantes sobre si mesmas e sobre a comunidade a que pertencem, tornando a internet uma grande fonte de informações não oficiais para saúde pública. Por isso, CHUNARA, ANDREWS e BROWNSTEIN (2012) apontam a importância da inovação dos sistemas tradicionais de vigilância em saúde através do uso de tecnologias como a internet e mostram que com o simples uso de ferramentas como as Mídias Sociais, a aplicação de clusters de palavras-chave, relacionadas a uma doença, tem o potencial de fornecer informações para a detecção precoce de eventos de saúde pública.

1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho apresenta a seguinte estrutura: no capítulo 1 que se encerra com esta seção, compreende o detalhamento da motivação, o problema, objetivo pretendido, a relevância da pesquisa e a organização do trabalho.

No capítulo 2, será apresentado o referencial teórico sobre o processo de identificação de rumores em saúde pública no Brasil, a evolução nas abordagens da vigilância em saúde e o uso de informações de Mídias Sociais em Saúde Pública.

No capítulo 3, serão descritos detalhes do método de pesquisa utilizado.

No capítulo 4, são descritos detalhes sobre a primeira entrevista não estruturada realizada com os especialistas na área de saúde.

O capítulo 5, apresenta Framework proposto.

O capítulo 6, apresenta uma avaliação conceitual feita pelos especialistas ao Framework e os resultados obtidos com duas instâncias do Framework.

O capítulo 7, apresenta a discussão da pesquisa e a possibilidade de sua evolução.

O capítulo 8, descreve as conclusões do trabalho.

2 REVISÃO DE LITERATURA

Neste capítulo são apresentadas as principais pesquisas associadas ao processo de identificação de rumores em saúde pública no Brasil, a evolução nas abordagens da vigilância em saúde e o uso de informações de Mídias Sociais em Saúde Pública.

2.1 RESPOSTAS ÀS EMERGÊNCIAS EM SAÚDE PÚBLICA NO BRASIL

A ocorrência de diversos casos de doenças, mudanças no cenário epidemiológico e o temor de ocorrência de novas crises globais levaram a Organização Mundial da Saúde (OMS), a revisar o Regulamento Sanitário Internacional (BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005). Com isto, esta agência de âmbito internacional que tem a responsabilidade de influenciar, monitorar e avaliar as políticas de saúde em todo o mundo através de suas unidades regionais, propôs novos conceitos e estratégias para os padrões de vigilância em saúde durante a realização da Assembleia Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2005).

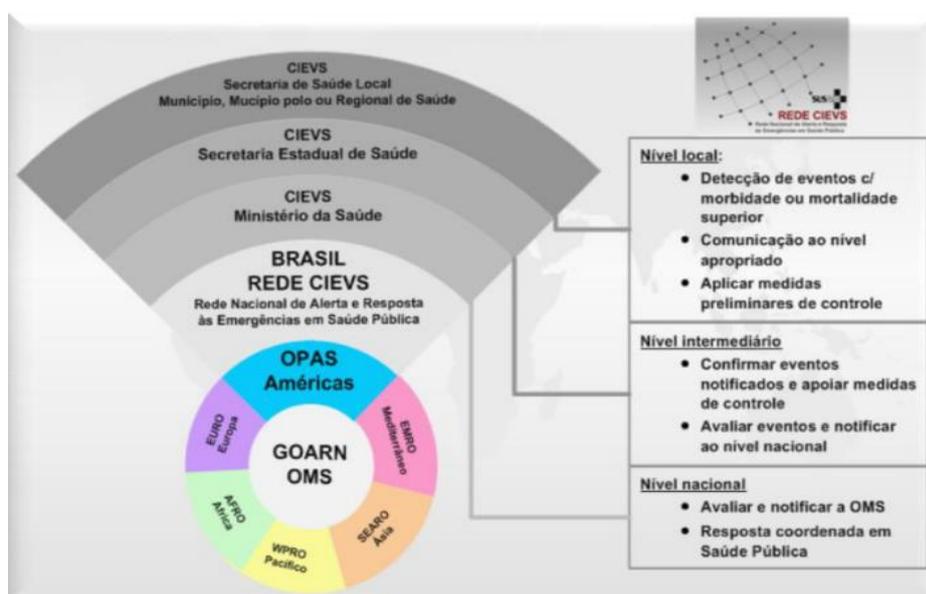
Com a realização desta assembleia, responsável pela revisão do RSI, diversos subsídios como políticas e padrões foram revistos ou instaurados para ampliar a capacidade de detecção e resposta aos riscos de disseminação de doenças e agravos. Com isto, a vigilância em saúde inicia seu monitoramento sobre determinados cenários logo em seu início, quando ainda são considerados riscos e não casos de emergência já concretizados o que possibilita uma atuação mais preventiva (BRASIL; SECRETARIA DE VIGILÂNCIA EM SAÚDE; CIEVS, 2018). Ao definir o conceito de evento na área da saúde pública (ESP) como sendo uma “manifestação de uma doença ou de uma ocorrência que cria um potencial para a doença”, o RSI de 2005 estabelece aos países membros da OMS que eventos em saúde pública não são apenas casos de doenças ou óbitos causados por ela, sua compreensão aborda também os possíveis riscos de sua ocorrência e as ações para evitar ou mitigar a probabilidade de uma propagação (BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005; TEIXEIRA; COSTA; CARMO; *et al.*, 2018). Com esta definição, o novo RSI amplia o escopo do monitoramento e intervenção da vigilância em saúde, incluindo consequências de desastres ambientais (p.ex. terremotos, inundações ou secas) e eventos de natureza química, biológica e rádio nuclear (BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE

VIGILÂNCIA EM SAÚDE AMBIENTAL E SAÚDE DO TRABALHADOR. *et al.*, 2014; BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005; FERGUSON *et al.*, 2005; TEIXEIRA; COSTA; CARMO; *et al.*, 2018).

Em 2006, como parte do processo de estruturação e aperfeiçoamento do serviço, recebimento, processamento e resposta às emergências epidemiológicas sugeridas pelo novo RSI (BRASIL; SECRETARIA DE VIGILÂNCIA EM SAÚDE; CIEVS, 2018), a Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS) instaurou os Centros de Informações Estratégicas em Vigilância em Saúde (CIEVS) que receberam a designação de Ponto Focal Nacional de contato permanentemente acessível à OMS. Além desta designação, os Centros possuem integração com a Rede Mundial de Alerta e Resposta às Emergência em Saúde Pública (Global Outbreak Alert and Response Network – GOARN), e se organizam em diversas unidades ativas em todo o Brasil que compõem a Rede Nacional de Alerta e Resposta às Emergências em Saúde Pública. Esta rede nacional tem como objetivo principal estabelecer a resposta coordenada a todo tipo de emergência em saúde pública, como surtos em comunidades, epidemias nacionais e internacionais, desastres e outras emergências em saúde pública (MINISTÉRIO DA SAÚDE *et al.*, 2015). Através dos serviços prestados por esses centros é possível a realização de monitoramento e tomadas de decisões necessárias para realização de respostas às possíveis emergências de saúde pública cada vez melhores (BRASIL. MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE VIGILÂNCIA EM SAÚDE AMBIENTAL E SAÚDE DO TRABALHADOR. *et al.*, 2014).

A Figura 1 ilustra a estrutura da vigilância sanitária e resposta a epidemias no Brasil, conforme estabelecido no artigo 5º e no anexo 01 do RSI (2005). Além dos CIEVS da Secretaria de Vigilância em Saúde no âmbito nacional, criados em 2005, nos anos seguintes foram criados os centros nos âmbitos estaduais (SES) e municipais (SMS) (BRASIL. *et al.*, 2016).

Figura 1 - Estrutura CIEVS no Brasil



Fonte: (FERREIRA; CASTRO, 2011)

Para realização de suas atividades, os CIEVs monitoram tanto informações oficiais pertencentes ao atual sistema de vigilância sanitária do Brasil, quanto informações não oficiais (rumores) obtidas de maneira ativa ou passiva que irão auxiliar na detecção de eventos na saúde pública. A obtenção ativa de rumores, refere-se a busca por rumores e identificação daqueles que podem apontar a existência de possíveis riscos, já a obtenção passiva de rumores, refere-se às informações recebidas diretamente da população que utiliza os diferentes meios de comunicação disponibilizados pelos Centros (telefone, e-mail e sites) ou através de outros sistemas na área da saúde. (MINISTÉRIO DA SAÚDE *et al.*, 2015). Na modalidade passiva, os Centros consideram que ao disponibilizarem vários meios de comunicação à população, não só motiva a sua participação, como também diversifica e facilita a captura de informações a respeito de eventos na saúde pública. A participação popular possui um grande peso em suas ações, pois notificações imediatas, contribuem para uma investigação mais apurada. Com uma melhor análise de riscos e adoção de medidas de prevenção e controle em tempo hábil, algumas ações podem ser tomadas sem a necessidade do aguardo de resultados de análises laboratoriais (BRASIL. *et al.*, 2016). Já a maneira ativa ocorre através da realização de serviços como: o constante monitoramento de informações; detecção e verificação de rumores sobre potenciais Emergências em Saúde Pública de Importância Nacional (ESPIN) ou

Internacional (ESPII); realização de investigações e resposta proativas à possíveis surgimentos de surtos e epidemias (BRASIL. *et al.*, 2016); aperfeiçoamento dos mecanismos de triagem; verificação e análise das notificações; interação entre secretarias de saúde nas diferentes esferas de governo (MS, SES, SMS); apoio a formulação de planos de respostas a emergências e apoio às áreas técnicas (BRASIL. *et al.*, 2007).

Com o tempo, essas modalidades distintas vêm se aproximando e ficando cada vez mais integradas. O aumento do uso de tablets, smartphones, jornais eletrônicos, mídias sociais etc. facilitou o compartilhamento de informações da população (LEAL-NETO *et al.*, 2016). Através do uso destes eletrônicos, surge um novo canal digital repleto de informações populares compartilhadas que auxiliam os canais tradicionalmente utilizados e auxiliam a atuação de profissionais de saúde no âmbito operacional (campo) comunicação (BEZERRA; BITOUN, 2017). Essa integração beneficia os CIEVS, que passam a ter informações mais rápida. Isto também estimula a contribuição popular e melhora a capacidade de responder ou mesmo prever as emergências na área da saúde (MINISTÉRIO DA SAÚDE *et al.*, 2015).

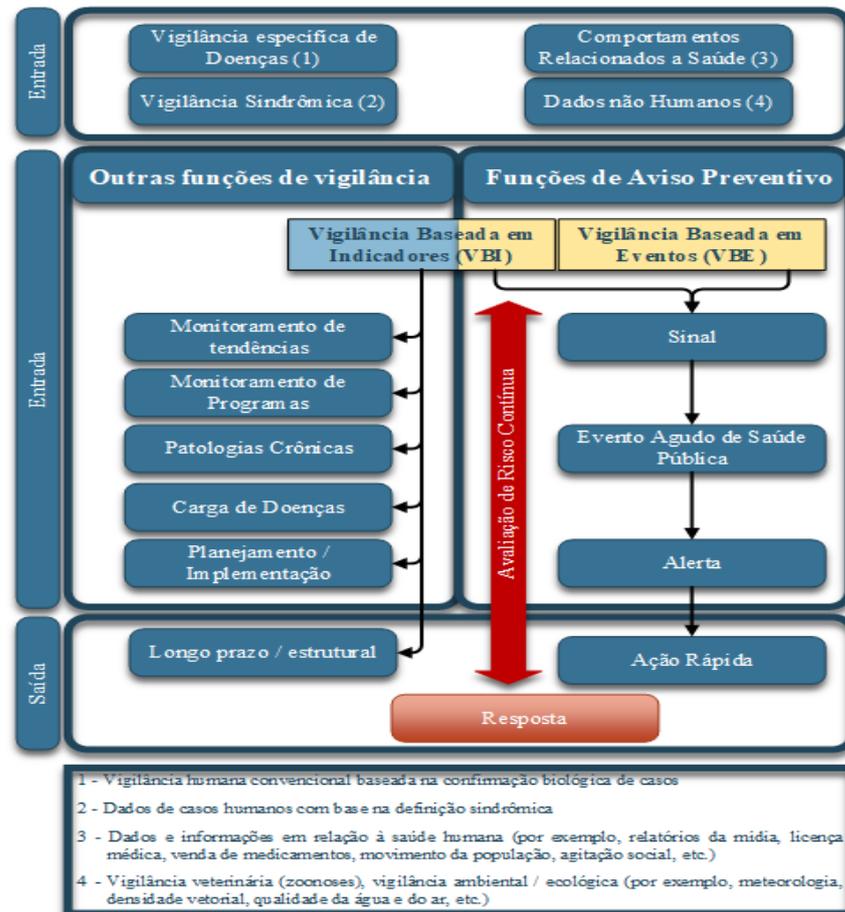
2.2 DA VIGILÂNCIA EM SAÚDE ATÉ A INTELIGÊNCIA EPIDÊMICA

Antes da revisão do regulamento a vigilância baseava-se no acompanhamento de relatórios e indicadores para parametrizar não só ações de resposta eficiente a possíveis emergências, mas também as ações voltadas aos programas sanitários. Também conhecida como Vigilância Baseada em Indicadores (VBI), esse processo utiliza informações que embora sejam de extrema importância, muitas vezes surgem tardiamente, incompletas ou limitam-se a abordagem de poucos riscos conhecidos à saúde pública, características geralmente ocasionadas pelo demorado processo de validação das informações, burocracias, alto custo ou mesmo necessidade de recursos. Segundo ZHOU,; YE e FENG (2011) na Vigilância Baseada em Indicadores, o aguardo da divulgação oficial de informações pode gerar redundâncias prejudiciais para fins de detecção precoce de possíveis surtos e consequente resposta ao mesmo.

Contudo, após a revisão do RSI em 2005, que passou a considerar novas fontes de informações para coleta e realização das análises da vigilância em saúde, surge uma nova abordagem, a Vigilância Baseada em Eventos (VBE), que tem o objetivo de elevar o nível

de detecção e resposta rápida a sinais e alertas advindos não só das fontes de informações oficiais, mas também das não oficiais (Figura 2).

Figura 2 - Visão Geral de Todas as Funções de Vigilância e Resposta à Saúde Pública



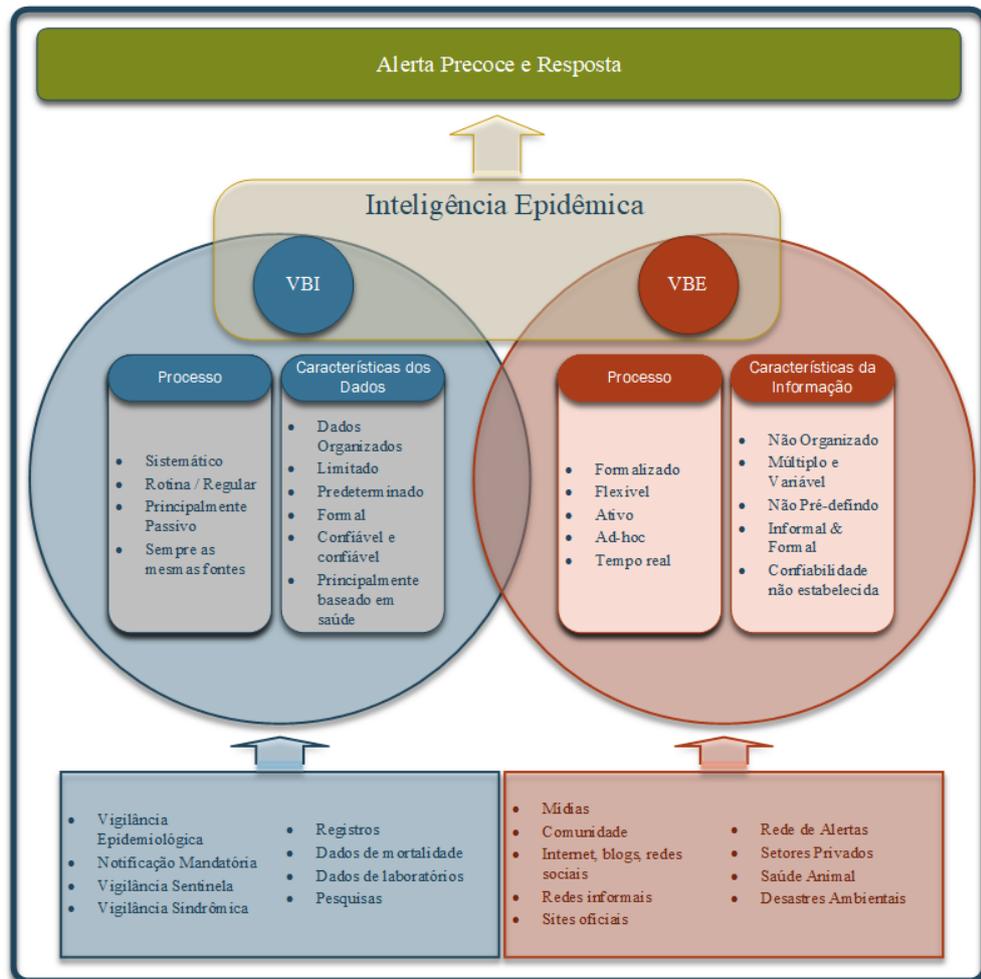
Fonte: (WORLD HEALTH ORGANIZATION, 2014)

Através da integração de ambos os tipos de vigilância (VBI e VBE), surge o conceito de Inteligência Epidêmica (IE) que fornece uma nova abordagem para enfrentar os desafios do dinamismo epidemiológico atual. A IE contempla as atividades realizadas para realização de uma resposta eficiente a um evento em saúde pública, conforme observado na Figura 3.

Considerada por BOWSHER, MILNER e SULLIVAN (2016) como um processo, ou mesmo um produto, capaz de analisar informações sobre vigilância de saúde para criação de percepções, a IE compreende a coleta, análise e comunicação de qualquer informação para detecção precoce, verificação e investigação de eventos e possíveis

riscos à saúde com objetivo de uma resposta antecipada a quaisquer possíveis emergências em saúde pública.

Figura 3 - Vigilância Baseada em Indicadores, Baseado em Eventos e Inteligência Epidêmica



Fonte: (WORLD HEALTH ORGANIZATION, 2014)

Através do novo RSI em 2005, a OMS concordou com a utilização de fontes de informações não oficiais nas atividades de vigilância em saúde. O uso de fontes de informações não oficiais permitiu que seus estados membros pudessem adequar seus sistemas de vigilância para utilizar este novo recurso. A partir de então, é permitido o uso de fontes de informações não oficiais como apoio ao processo tradicional, para complementar sua contextualização do cenário e permitir um resultado mais preciso. (BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005).

No Brasil, através de atividades como a Detecção Digital de Doenças (DDD) e a Vigilância Participativa, os CIEVS buscam informações não oficiais (rumores) referentes a sinais e sintomas de doenças tanto em meio digital quanto em meio a relatos realizados por indivíduos de uma comunidade (BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2005). Através da estratégia de Vigilância participativa, a população é envolvida no processo, tornando os indivíduos uma fonte primária de informação não oficial (coletada diretamente de usuários) e pode contribuir com relatos de informações independentes dos sistemas de saúde utilizados auxiliando na contextualização dos cenários epidemiológicos (GREENLEAF *et al.*, 2021). Com a estratégia de DDD, a vigilância em saúde busca utilizar a tecnologia da informação como apoio aos processos de vigilância para analisar diversas fontes de informações na internet e compreender um determinado cenário epidemiológico (DENECKE, 2017).

As Mídias Sociais têm se mostrado bastante úteis, pois através delas as pessoas compartilham e discutem suas opiniões ou experiências e muitos compartilham suas informações relacionadas à saúde (PAUL *et al.*, 2016). Dada a quantidade de dados e os diversos estilos e palavras usadas pelas pessoas para se referir a um tópico, a pesquisa e coleta de postagens também é um trabalho desafiador (VEYSEH *et al.*, 2017). Na tentativa de superar este desafio de identificação de informações relevantes nas Mídias Sociais por meio de monitoramento diário, os CIEVS utilizam alguns agregadores de postagens, como o Tweetdeck.

O Tweetdeck reúne as informações, segundo alguns critérios estabelecidos (hashtags), e os organiza em colunas distintas. Isto permite aos usuários visualizarem múltiplos tweets em tempo real e simultaneamente (DAKKAK *et al.*, 2018). Porém, essas mensagens não são armazenadas para realização de uma análise mais criteriosa pelas equipes, possibilitando que informações importantes possam passar pelo monitoramento sem serem percebidas pelas equipes do CIEVS.

Com essas estratégias, os centros atribuem informações adicionais ao processo tradicional de monitoramento e vigilância em saúde, implementando os conceitos sugeridos pela Inteligência Epidêmica, em busca de uma detecção precoce de eventos em saúde pública, sua verificação e uma rápida resposta a possíveis emergências.

2.3 MÍDIAS SOCIAIS NA VIGILÂNCIA EM SAÚDE

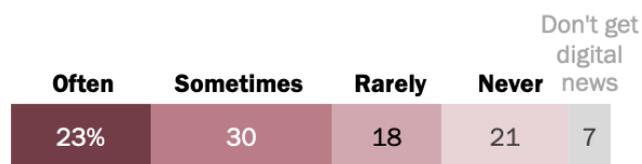
A internet disponibilizou aos usuários, um espaço dinâmico onde as informações são constantemente produzidas, compartilhadas e consumidas (GIUSTINI, 2006). Paralelamente ao seu crescimento, esse novo ambiente trouxe consigo um grande aumento do volume de informações disponíveis e capazes de serem usadas para as mais diversas finalidades (DODGE, 2017; VERMA *et al.*, 2011).

Entre todas as ferramentas utilizadas para realização de buscas ou compartilhamento de informações em uma discussão, as Mídias Sociais têm ganhado destaque e maior aceitação. Segundo MITCHELL (2021), mais da metade dos usuários das Mídias Sociais adultos e americanos, relataram o recebimento de notícias com determinada frequência, conforme exibido no Gráfico 1. Isto faz com que as Mídias Sociais tenham maior destaque perante a população como fonte de informações críticas ou mesmo de notícias diversas, incluindo aquelas relacionadas à saúde (HAGEN *et al.*, 2020).

Gráfico 1 - Frequência de Busca por Informações nas Mídias Sociais

About half of Americans get news on social media at least sometimes

% of U.S. adults who get news from social media ...



Note: This chart is not comparable to similar questions asked in the past due to question wording changes; see Appendix for more details.

Source: Survey of U.S. adults conducted Aug. 31-Sept. 7, 2020. "News Use Across Social Media Platforms in 2020"

PEW RESEARCH CENTER

Fonte: (MITCHELL, 2021)

DODGE (2017) também aponta que as Mídias Sociais têm se destacado como novo meio de comunicação, modificando a forma como este processo é realizado atualmente. Ao longo dos anos, através de seus serviços online e móveis, as Mídias Sociais estimulam e viabilizam um maior compartilhamento de informações pelos seus usuários que falam sobre si próprio ou sua comunidade (KAPLAN; HAENLEIN, 2010). Este compartilhamento constante, gerou um maior volume de informações disponíveis que abordam diversos contextos (opinião pública, questões nacionais, testemunho ocular etc.), podendo causar uma sobrecarga de informações e dificultar a identificação daquelas que são relevantes. Para superar essa dificuldade, alguns pesquisadores fizeram uso de importantes recursos disponibilizados pelas Mídias Sociais, os metadados das postagens. Estes recursos representam informações adicionais que acompanham cada postagem e são capazes de descrever melhor o contexto em que ela foi produzida (data/hora, localização etc.) (PAUL; DREDZE, 2017; PEREZ; MUSOLESI; STRINGHINI, 2018). Essa estratégia permite melhorar a exploração das informações de Mídias Sociais para pesquisas em muitas áreas, incluindo a área da saúde (LIAO *et al.*, 2020; MOORHEAD *et al.*, 2013).

Dentro da área de vigilância da saúde pública, as mídias sociais tornaram-se grandes ferramentas para detectar, monitorar, prever tendências epidêmicas ou mesmo prestar informações ao público. Mesmo não sendo projetadas com esta finalidade específica, o seu elevado volume de informações e respectivos metadados, auxiliam na complementação da infraestrutura tradicional de vigilância, que geralmente disponibiliza informações após um atraso significativo (PLATT; CITRIN; HOOD, 2015). Esse atraso pode ocasionar dados redundantes ou mesmo dados obsoletos para uso em ações de prevenção (ZHOU; YE; FENG, 2011).

Em 2017, JONES *et al.* observaram que as pessoas costumam buscar e fornecer informações relacionadas a eventos em saúde para se manter informado. No entanto, quando há falta de informações advindas dos canais oficiais, estas pessoas estão suscetíveis a rumores que geralmente preenchem o vazio de algumas informações. Esses rumores são identificados como uma declaração onde a veracidade das informações não foi verificada, diferentemente das informações advindas de fontes oficiais (DIFONZO; BORDIA, 2007) e se originam em meios de comunicação sociais, como as Mídias Sociais (OZTURK; LI; SAKAMOTO, 2015).

Considerando que publicações sobre o estado de saúde do próprio usuário são comuns nas Mídias Sociais. Afirmações como “eu tenho gripe”, quando analisada

isoladamente pode não ter grande significado, mas a análise de diversas postagens com conteúdo semelhante, pode trazer um novo significado que permita um possível rastreamento sobre rumores relacionados a síndromes. Pensando nisso, diversos pesquisadores elaboraram estudos para conseguir acompanhar ou mesmo detectar antecipadamente surgimento de surtos de doenças através da análise dessas informações compartilhadas pelas Mídias Sociais (CHEW; EYSENBACH, 2010; DIAZ-AVILES; STEWART, 2012; WAKAMIYA; KAWAI; ARAMAKI, 2018).

Entre todas as plataformas de Mídias Sociais, SIGNORINI (2009) destaca o Twitter como um importante meio de comunicação em todo o mundo. Segundo o autor, a possibilidade de utilização do Twitter através de diversos dispositivos possibilita a realização de uma publicação de maneira fácil e rápida. Em sua pesquisa, o autor aponta recursos do Twitter que auxiliam a extração de informações úteis e capazes de caracterizar a existência de rumores em saúde e ajudam a evitar o efeito de sobrecarga de informações.

Então, filtrar ou mesmo reestruturar as informações de Mídias Sociais de maneira que facilite a extração daquelas que sejam úteis e descartando possíveis “ruídos” desnecessários é de suma importância. Isso vem sendo demonstrado através das pesquisas realizadas ao longo dos anos, onde os autores estudam as melhores maneiras de utilizar o potencial existente das informações de Mídias Sociais.

Em 2010, os pesquisadores CHEW e EYSENBACH sugeriram o uso de informações oriundas do Twitter para complementar a vigilância de saúde tradicional. Eles analisaram a ocorrência de mensagens que continham as palavras-chave “H1N1” e “Gripe Suína” para validar o Twitter como uma ferramenta, não só para analisar o comportamento de uma doença, mas também para acompanhar outras informações sobre o público em tempo real. A pesquisa analisou mais de 2 milhões de tweets capturados entre 01 de maio a 31 de dezembro de 2009 e observou que mensagens relacionadas ao termo “H1N1” foram utilizadas principalmente para disseminar informações de fontes confiáveis para o público e serviram de fonte de opiniões e experiências dos usuários. Os autores ressaltaram que esse conteúdo pode ser usado em tempo real para diversas finalidades, incluindo análise de sentimento do público. A pesquisa não demonstrou uma estrutura para o estabelecimento de palavras-chave para captura das mensagens, por isso os resultados podem ter deixado passar qualquer mensagem útil que não contivesse a palavra “H1N1” ou que a utilizasse com qualquer outro sinônimo.

No ano seguinte, HU, TANG e LIU (2011) observaram que Micro blogs como o Twitter que possuem textos curtos, são capazes de disponibilizar tanto informações úteis

quanto meros ruídos. Segundo eles, as informações úteis em meio a tantas outras não estruturadas são de difícil identificação e acesso. Como solução, os autores propuseram o agrupamento dessas mensagens em clusters com atribuição de rótulos semanticamente significativos e a aplicação de uma nova estrutura para organizar estas informações não estruturadas em uma estrutura semanticamente organizada. Técnicas de Processamento de Linguagem Natural (PLN) foram aplicadas para realização de uma decomposição sintática e posterior mapeamento semântico que auxiliasse na compreensão das mensagens. Em seguida, os autores estabeleceram clusters e rótulos para facilitar a organização do grande número de informações e agrupá-las segundo suas características semânticas. Após a análise de 11.362 tweets, a pesquisa apresenta resultados satisfatórios em relação a outros métodos de identificação de mensagens úteis em Micro blogs, porém não mostra preocupação na identificação e descarte de mensagens não úteis de maneira estruturada.

Em 2012, a pesquisa realizada por DIAZ-AVILES, e STEWART estudou as mensagens do Twitter durante o surto de *Escherichia coli* enterohemorrágica (EHEC) na Alemanha, com intuito de entender melhor o comportamento populacional neste período e analisar a capacidade de emissão de um alerta em saúde de maneira preventiva. Identificada frequentemente como causadora da síndrome hemolítico-urêmica (SHU), gastroenterite (GI) e diarreia hemorrágica, a EHEC teve a contabilização de milhares de pessoas infectadas no país. Através do apoio do Instituto Robert Koch (IRK), autoridade de saúde nacional, a pesquisa monitorou mais de 500 doenças e sintomas junto ao Twitter. Durante o período de maio a junho de 2011 foram coletados 7.710.231 tweets relacionados a condições médicas e 456.226 relacionados diretamente ao surto EHEC, originadas através dos casos relatados pelo IRK e uso da palavra-chave “EHEC” respectivamente. A pesquisa analisou as mensagens, reconheceu a potencial contribuição das informações do Twitter, porém identificou a dificuldade em analisar o grande volume de informações obtidas e a necessidade de mecanismos capazes de ajudar o processo.

Em 2014. SANTOS e MATOS classificaram manualmente mais de 2000 tweets e logs de pesquisa de internet de Portugal para treinar um classificador capaz de identificar tweets que mencionam gripe ou doenças semelhantes. Através do uso deste classificador, a pesquisa propõe um modelo preditivo de casos de influenza em Portugal. Para captura destas mensagens, a pesquisa utilizou apenas a palavra “gripe” e algumas de suas derivações, não houve relação com outros sintomas da doença e seus respectivos sinônimos. Para filtragem dos tweets capturados, foram utilizadas outras palavras-chave

com suas respectivas derivações. Após todo esse processo, os 3.183 tweets foram classificados para contribuir e testar o modelo proposto. A aplicação de um processo de refinamento dos parâmetros aplicados na análise, poderia apresentar melhores resultados para aplicação e melhorar o modelo proposto pela pesquisa.

No ano seguinte, em 2015, JAIN e KUMAR tentaram rastrear uma pandemia de H1N1 na Índia através da utilização do Twitter. A pesquisa utilizou palavras-chave relacionadas à doença que foram consideradas frequentes nos principais jornais. A busca por tweets relevantes foi realizada no período de 15 a 28 de fevereiro de 2015, chegando a um montante de 91495 mensagens. A análise destas mensagens se deu através da aplicação de técnicas de classificação para identificar as mensagens relevantes e as irrelevantes. Os ajustes dos classificadores utilizados foram baseados em amostras das mensagens sem avaliar todas as mensagens obtidas. Através do estabelecimento de estratégias de refinamento da análise e abordagem de todas as informações coletadas, a pesquisa poderia ter mitigado o descarte indevido de informações relevantes.

Mais tarde, em 2020, BAKER *et al.*, utilizaram técnicas de aprendizado de máquina sobre tweets em árabe para detectar a existência de surto de gripe em países árabes. Com uso de 34 palavras-chave relacionadas à gripe, a pesquisa coletou 54.065 tweets, que após uma análise manual, reduziu o montante a 1473 mensagens capazes de identificar um surto de gripe nos países pesquisados. A pesquisa não menciona nenhuma abordagem a variações ou sinônimos das palavras-chave utilizadas na coleta para aumentar a abrangência, nem estabelece um critério de análise mais robusto para identificar suas mensagens relevantes e descartar aquelas com pouca ou nenhuma contribuição.

Em 2017, VEYSEH *et al.* analisaram 100.000 tweets para entender o posicionamento de usuários do Twitter em relação a publicação de boato propagado por um tweet. Considerando que ao identificar um boato, o usuário pode apoiar, negar, consultar ou comentar a respeito, os pesquisadores propuseram a utilização de um modelo capaz de analisar não só esse tweet identificado, mas também aqueles que são vizinhos. A aplicação do modelo resulta na postura do usuário quanto aquele boato com maior precisão do que outros métodos usados anteriormente. Esta pesquisa estudou outro aspecto importante das Mídias Sociais, o compartilhamento das informações. Porém, não demonstrou uma estruturação organizada para as informações analisadas.

No ano seguinte, o estudo realizado por WAKAMIYA, KAWAI e ARAMAKI (2018) propôs um modelo capaz de utilizar as mensagens do Twitter para detectar a existência de possíveis eventos em saúde pública. O modelo foi avaliado com

informações relacionadas à influenza, coletadas durante 3 anos seguidos. O modelo divide as informações úteis em direta, quando o autor menciona seu próprio estado, ou indireta, quando o autor menciona algo sobre outra pessoa. Desta maneira, o modelo espera compensar o desequilíbrio na distribuição geográfica da população japonesa que utiliza o Twitter. O modelo aplica técnicas de Processamento de Linguagem Natural (PLN) sobre as mensagens para identificar aquelas que podem sugerir a existência de casos de influenza no Japão. Ao final da pesquisa, os autores concluíram que as informações diretas têm maior contribuição em locais mais urbanos, detentores de maior volume de usuários do Twitter e as informações indiretas têm maior contribuição em áreas mais remotas por terem poucos usuários. Ainda segundo os autores, a pesquisa teve um fator limitador, a utilização de apenas uma palavra-chave para captura das mensagens. Segundo eles, o uso de apenas uma palavra-chave específica, restringe a captura das informações e inviabiliza o estudo de outros pontos de vista do cenário, como ordem cronológica. Um conjunto de palavras-chave relacionadas aos sintomas da doença sob investigação e seus sinônimos, poderiam ter superado essa limitação e trazido mensagens com maiores contribuições para a investigação em andamento. A pesquisa também não exibe um tratamento mais estruturado das mensagens encontradas junto ao Twitter.

Pouco tempo depois, VILAIN, MENUJER e FILLEUL (2019) coletaram 238.244 tweets no idioma francês que continham os termos “gripe”, ”gripal”, ”grippaux”, coletadas entre 08/08/2016 e 26/04/2017, para avaliar se o Twitter pode ser uma fonte de dados complementar para monitorar a gripe. Destas mensagens, os autores descartaram todas que não estivessem relacionadas à França, o que reduziu as informações para um total de 130.559 tweets. Após uma etapa de limpeza, as mensagens foram reduzidas para 22.939 tweets. Após a classificação dessas mensagens por um classificador previamente treinado, a pesquisa observou que a cronologia dessas mensagens é semelhante a cronologia de registro de visitas ao pronto-socorro e consultas médicas para influenza, confirmando a utilidade do Twitter como fonte de informações sobre saúde. As palavras-chave utilizadas não abordaram os sintomas da doença sob investigação, muito menos seus sinônimos. As mensagens foram classificadas uma única vez, sem aplicação de refinamentos ou uma fase distinta do processo para uma análise mais aprofundada do conteúdo da mensagem.

No mesmo ano, LIANG *et al.* (2019), pesquisaram sobre padrões de difusão de informações sobre o Ebola no Twitter para identificar possíveis usuários influentes, capazes de auxiliar na comunicação de saúde pública durante a resposta a emergências. A pesquisa coletou 36.931.362 tweets compreendidos no período de 23 de março de 2014

a 31 de maio de 2015 para analisar seus conteúdos textuais, autores e seus seguidores. Mesmo com uso de algumas palavras-chave e hashtags para captura da mensagem, a pesquisa não demonstrou uma análise posterior mais apurada, com uso de novos parâmetros, que pudessem identificar mensagens mais relevantes para a pesquisa. Então, o estudo limitou-se a uma amostra de 192.209 tweets de todas as informações obtidas. Após a análise dessa amostra, a pesquisa identificou a existência de dois modelos de comunicação no Twitter: a transmissão e o viral. Posteriormente, através desses modelos, a pesquisa identificou usuários influentes para disseminação de dados do Twitter referentes ao Ebola. A classificação proposta para os usuários, não se baseou no conteúdo textual publicado no tweet e sim em características como quantidade de retweets da mensagem de um usuário e sua quantidade de seguidores. A aplicação de uma análise refinada, poderia identificar se o usuário considerado influente é também um produtor de mensagens relevantes para a comunicação em saúde pública ou se é um produtor de mensagens irrelevantes que pode ter sua mensagem descartada através de um processo adequado.

Ainda neste mesmo ano, MONDAL *et al.* (2019) pesquisaram sobre detecção, monitoramento e mitigação de propagação dos rumores nas Mídias Sociais. Para isso, eles coletaram e analisaram mais de 1 milhão de tweets em meio a um período de chuvas fortes na cidade de Chennai em 2015. As mensagens foram coletadas de 01 a 10 de dezembro de 2015 e tiveram sua natureza analisada sobre os aspectos temporais, estruturais e linguísticos, observando a existência de mudanças em decorrência de um desastre. A pesquisa estabeleceu 4 clusters, sendo dois referentes a rumores e dois referentes a não rumores, e em seguida analisaram cada mensagem em busca de identificar a sua relação com algum dos “clusters”. Caso a mensagem não mostrasse relação a qualquer um dos “clusters”, ela era descartada. Todo o processo de análise foi realizado por dois participantes da pesquisa, não apresentando a utilização de uma estrutura ou automação para análise das mensagens.

Outra pesquisa em 2019 feita por DUVAL e SILVA desenvolveu um sistema para utilizar informações do Twitter na área de farmacovigilância. Segundo os autores, Eventos Adversos (EA) são agravos à saúde que podem surgir durante o tratamento com um produto farmacêutico, podendo ser erro de medicação, desvio de qualidade dos medicamentos ou reações adversas a medicamentos (RAM) que no seu período de pós-comercialização requerem maiores cuidados com a saúde para evitar danos desnecessários ao paciente. A pesquisa utilizou o Twitter como fonte para mineração de

texto para identificar mensagens relacionadas a eventos adversos relacionados à doxiciclina, medicamento de tratamento da malária. Durante todo o ano de 2014 e através do uso da nomenclatura de 19 medicamentos no combate à Malária, foram capturadas 6,9 milhões de mensagens que demonstrou ter a doxiciclina com maior número de mensagens relacionadas. Em seguida, com uso de uma ontologia voltada para identificação de EA, identificou-se 4.401 mensagens relacionadas ao medicamento que tiveram seu conteúdo analisado e comparado com registros do Food and Drug Administration (FDA). A pesquisa fez a coleta de um grande volume de informações através da utilização de poucos parâmetros, o que se expandiu pelo período de um ano. A pesquisa não descreve uma análise das mensagens de maneira mais estruturada, que possa reduzir ainda mais as mensagens relevantes relacionadas ao medicamento doxiciclina e não permite nenhum processo de refinamento e reutilização da análise sobre as mensagens na busca de melhores resultados.

Mais recentemente, em 2020, a pesquisa de LWIN *et al.* também utilizou as informações de Mídias Sociais. Através das mensagens do Twitter, a pesquisa estudou as tendências mundiais de quatro emoções: medo, raiva, tristeza e alegria durante a pandemia da COVID-19. Através do uso das palavras-chave (Wuhan, corona, nCov e Covid), foram identificadas mais de 20 milhões de postagens no período de 28 de janeiro a 9 de abril de 2020. Através da aplicação de uma tecnologia para análise de sentimento, a pesquisa analisou as tendências das emoções ao longo do tempo. A pesquisa sugere que questões coletivas movidas pelas emoções relacionadas à pandemia estavam associadas ao isolamento social, a perda de vidas humanas e aponta a necessidade de monitoramento e controle das preocupações sociais para manutenção do bem-estar psicológico público. A utilização de poucas palavras-chaves não relacionadas aos sintomas da COVID-19 permitiu a identificação de um grande volume de mensagens para realização da análise de sentimentos. A utilização de termos mais estruturados na captura permitiria a identificação de mensagens com maior relação com a COVID, viabilizando até mesmo a identificação de pessoas com COVID e a análise de suas emoções.

No mesmo ano, outra pesquisa buscou informações do Twitter para auxiliar a vigilância em saúde no monitoramento de uma epidemia de opioides (OODS) na Carolina do Norte nos Estados Unidos entre os anos de 2009 e 2017. Em sua pesquisa, ANWAR *et al.* (2020) reuniu um grupo de palavras-chaves que considerou serem aptas a identificarem informações úteis junto ao Twitter e obteve como resposta um grande volume de mensagens, mais de 100.000. Uma amostra aleatória de 10.000 mensagens

teve seu conteúdo analisado na íntegra, sem utilização de uma fase específica, com estrutura definida. A exclusão das mensagens também não demonstrou qualquer estruturação aplicada, foram excluídas todas as mensagens que contivessem links, sem antes realizar um tratamento do conteúdo. Sobre as mensagens restantes, a pesquisa prosseguiu com suas análises e aplicações estatísticas e concluiu que o Twitter é uma fonte de informações para monitoramento da epidemia de opioides.

Ainda em 2020, HAGEN *et al.* exploraram o potencial não só de analisar informações oriundas do público, mas também de como fornecer informações oficiais a ele. Eles analisaram o conteúdo do Twitter em meio a uma crise do Vírus Zika no sudeste dos Estados Unidos e observaram o nível de aderência e utilização do governo para realização de comunicação, não só sobre a crise, mas também sobre possíveis riscos existentes. A pesquisa coletou tweets contendo apenas a palavra-chave “Zika” durante o período de 25 de agosto a 21 de setembro de 2016. As análises realizadas foram aplicadas sobre mensagens selecionadas aleatoriamente após a coleta. Um dos resultados foi a identificação de que mesmo as contas de órgãos federais possuindo três vezes mais seguidores do que os órgãos de outras esferas, esses demais órgãos reunidos fazem maior uso de tweets, menções e retweets. Por não estabelecer um processo estruturado tanto para captura quanto para análise das mensagens, a pesquisa teve suas análises baseadas em amostragens aleatórias. Com esta estratégia, mensagens importantes podem não ter sido capturadas ou mesmo descartadas.

No mesmo ano, FREITAS, BORGES e CARVALHO (2020), saíram um pouco da área da saúde e propuseram um framework para construção de soluções tecnológicas que contribuam para a utilização das informações de Mídias Sociais pelas equipes de resposta a emergências. O framework estabelece parâmetros para selecionar, classificar, priorizar e identificar as informações relevantes ao contexto de emergência. O framework proposto trabalha com informações em tempo real e permite que as equipes responsáveis possam gerar ações de resgate mais eficientes. Através da aplicação dos parâmetros do Framework sobre informações reais, a pesquisa demonstrou que entre diversas informações disponíveis, foi possível encontrar um pequeno percentual que realmente possuía potencial contribuição para as equipes em suas operações de resposta e resgate.

Ainda em 2020, GOMES PESSANHA *et al.* fizeram uma análise de sentimentos dos usuários do Twitter sobre a Covid-19. Durante o período de janeiro a julho de 2020, eles utilizaram as hashtags Covid-19, FiqueEmCasa e algumas variações para captura de 56.090 tweets. O estudo identificou uma relação entre as publicações e a disseminação da

doença pelo Brasil e identificou também que o sentimento negativo denota a preocupação da população em relação ao avanço da doença e suas consequências. A pesquisa realizou uma limpeza textual nas mensagens, não só para ajudar nas análises das mensagens, mas também para exclusão de algumas mensagens, cujo critério não foi detalhado. A análise das mensagens foi realizada de maneira a identificar as mensagens com características positivas, neutras ou negativas, através de uma análise de frequência de palavras específicas. O trabalho não demonstrou muito critério no estabelecimento das palavras-chave para captura das mensagens, atendo-se apenas a hashtags e não a sintomas voltados à doença. Esse tipo de abordagem poderia trazer informações mais relevantes, cujo sentimento do autor da mensagem ficasse mais bem retratado. Em seguida, a limpeza da mensagem já realizou exclusões, mas não demonstrou quantidade de descarte e nem critérios deste processo. A pesquisa identificou similaridade nos seus resultados com informações de outras fontes sobre a Covid-19, contudo a análise destas mensagens não possui critérios nem refinamentos para reduzir o volume de mensagens e identificar aquelas com maior contribuição, reduzindo o esforço de análise.

Mais recentemente, em 2021, SIMÕES e NEVES identificaram que no Brasil, os sistemas oficiais responsáveis por acompanhar os dados da pandemia de Covid-19 não são integrados. Então, os autores propuseram a realização de uma mineração de texto no Twitter para analisar a tendência epidemiológica da Covid-19 no país, ao longo do tempo, para mostrar a capacidade do Twitter em auxiliar o monitoramento de epidemias. A pesquisa ressalta que a falta de integração dos sistemas atuais de monitoramento, casos de duplicidade de dados e discrepâncias entre as datas lançadas no sistema e as datas reais, prejudicam o apoio prestados por estes sistemas aos sistemas de vigilância em saúde. Através do uso da palavra-chave covid19 com alguns de seus sinônimos (corona, #coronavirusPlantaio) e alguns sintomas da doença (falta de ar, febre e tosse) a pesquisa identificou 59 mil tweets. A limpeza das mensagens foi realizada em cima de uma amostra aleatória de 9.600 tweets que resultou em 7.179 mil mensagens tratadas e que foram manualmente classificadas. Após a classificação, a pesquisa apresenta seus resultados e constata que o Twitter é eficiente em auxiliar no monitoramento e prevenção de pandemias. Em seguida, a pesquisa extrai uma amostra aleatória para realizar suas análises, mas vale ressaltar que isso gera também o simples descarte de mensagens de maneira aleatória, o que possibilita o descarte de mensagens relevantes para a pesquisa. Neste momento, a existência de uma estrutura que organizasse a análise, descarte e uma possível reanálise, poderia apontar resultados mais promissores para a pesquisa e facilitar

qualquer necessidade de atividade manual, como foi a classificação manual de todas as 7.179 mil mensagens.

É importante ressaltar que durante o mapeamento do estado da arte sobre o uso das informações de Mídias Sociais no processo de vigilância em saúde, não foi possível identificar a utilização de uma estrutura que organizasse de maneira mais criteriosa todo o processo de captura e análise destas informações para que fossem utilizadas no processo de detecção e monitoramento de rumores em saúde pública. Ao analisar essas publicações, foi possível observar o potencial do Twitter não só para coleta de informações do público com suas opiniões e experiências, mas também o seu potencial como disseminador de informações de fontes confiáveis ao público. Foi possível observar também, a necessidade de maior atenção ao tratamento final de tais mensagens para seu posterior uso em processos decisórios.

3 MÉTODO DE PESQUISA

Definido por GIL (2008), método científico é “conjunto de procedimentos intelectuais e técnicos adotados para se atingir o conhecimento”. Portanto, é possível considerar que o objetivo da ciência é chegar à veracidade dos fatos por meio de um método que permita atingir um determinado conhecimento.

O método utilizado neste trabalho, iniciou-se através de uma pesquisa bibliográfica para mapear o estado da arte sobre o processo de resposta a emergências em saúde no Brasil, sua estrutura, sua regulamentação, suas abordagens (indicadores e eventos), quem são as equipes envolvidas e como é realizado o fluxo de comunicação, monitoramento e detecção de rumores. Através da realização desse mapeamento foi possível identificar as contribuições no meio científico, analisar os principais resultados, as lacunas abertas, as áreas ainda não exploradas (SILVA; CARVALHO, 2014) e a identificação da existência do problema, um importante passo para o sucesso de um trabalho científico. O problema observado está relacionado ao grande esforço necessário para estruturar as informações relevantes junto às Mídias Sociais, para o processo de detecção de rumores em saúde pública.

A partir do momento em que a pesquisa buscou informações sobre a atual maneira de utilização das informações de Mídias Sociais pelas equipes de detecção e rumores em saúde pública, foi identificado esse grande esforço que é realizado pelas equipes ao analisarem essas informações em busca de informações relevantes. O grande volume de informações disponíveis e o seu rápido crescimento, dificulta todo o processo de análise e possibilita que informações relevantes possam não serem identificadas.

Para mapear e compreender mais detalhadamente sobre como as informações de Mídias Sociais são utilizadas para detectar rumores em saúde pública, entender suas atitudes, valores e motivações (MINAYO, 2010), esta pesquisa realizou uma entrevista na modalidade não estruturada com o grupo de especialistas e profissionais nos processos de detecção de rumores realizados pelos CIEVS. A modalidade aplicada teve como propósito ampliar a captação das opiniões dos participantes (CRESWELL, 2012), elevando o nível de compreensão sobre todo o processo realizado pelos CIEVS e sua importância para realização de uma resposta oportuna a emergências em saúde pública. A partir do embasamento adquirido pela pesquisa bibliográfica e do entendimento advindo das entrevistas iniciais, foi observado a necessidade de uma solução capaz de

sugerir uma melhor estruturação das informações a serem usadas para o processo de detecção de rumores.

Entender a aderência dos conceitos abordados em pesquisas ligadas às tecnologias da informação e comunicação (TIC) é parte importante de uma pesquisa e pode ser auxiliada através da construção de artefatos capazes de resolverem problemas do mundo real (SIMON; LAIRD, 2019). Com isso, este trabalho propõe um Framework Conceitual, ou seja, “um projeto genérico em um domínio que pode ser adaptado para aplicações específicas, sendo útil como um molde para construção de aplicações” (SOUZA, 1998). Através do framework proposto, busca-se orientar o desenvolvimento de sistemas que sejam capazes de realçar os conceitos relevantes e suas interações dentro do domínio de vigilância em saúde pública em suas atividades de monitoramento e detecção de rumores em saúde pública. Concebido de forma a atingir o objetivo final proposto de maneira eficiente, o framework preserva a coesão das atividades abordadas através da definição de fases distintas que interagem entre si e agregam valores à fase posterior até o final do processo.

Após a construção da solução é importante avaliá-la (WIERINGA, 2009). Então, após a construção do framework, esta pesquisa apresentou a solução elaborada ao mesmo grupo de profissionais entrevistados anteriormente. Nesta apresentação foram demonstradas as fases do Framework, suas funcionalidades, seu escopo, utilidade e objetivos pretendidos. Ao final, foram coletadas as percepções e sugestões de melhorias.

Para mostrar o comportamento do Framework proposto ao ser aplicado em um cenário real, com utilização de informações reais coletadas do Twitter, esta pesquisa elaborou duas instâncias que representam duas especializações distintas dentro do domínio de monitoramento e detecção de rumores em saúde pública (GREGOR; HEVNER, 2013; SEIN *et al.*, 2011). Cada uma das instâncias, será aplicada com escopo e localidades distintas, no intuito de apresentar a aplicabilidade do framework. Sobre as informações obtidas em cada uma das situações, serão aplicados os passos sugeridos pelo framework no intuito de verificar a existência de um rumor que aponte a necessidade de emissão de alerta para uma investigação em cada uma das localidades.

Os resultados obtidos foram apresentados e discutidos com as equipes de profissionais anteriormente entrevistados para validação dos resultados obtidos e a sua contribuição para responder à questão abordada por esta pesquisa.

Através do processo metodológico exibido na Figura 4, o framework terá sua aplicação validada no intuito de responder à pergunta de pesquisa sobre a estruturação

das informações das Mídias Sociais para facilitar o seu uso pelas equipes de detecção de rumores em saúde pública.

Figura 4 - Método da Pesquisa



4 ENTREVISTA NÃO ESTRUTURADA

Uma entrevista não estruturada foi realizada com um grupo de especialistas e profissionais do CIEVS com intuito de entender melhor o processo de monitoramento e detecção de rumores realizado por estas equipes. A entrevista foi composta principalmente por questões abertas, no intuito de deixar os entrevistados falarem livremente (CAREGNATO, 2017).

A Tabela 1 exibe o questionário realizado e o resumo da resposta dos especialistas

Tabela 1: Questionário da entrevista

QUESTÕES	PROFISSIONAL 1	PROFISSIONAL 2	PROFISSIONAL 3
Tempo experiência (Anos)	3 anos	1 ano	9 anos
Como é realizada a busca por rumores tradicionalmente?	É realizada através do uso de outras ferramentas e agregadores de notícias. Agregadas as notícias de diversas fontes, eu observo o conteúdo para detectar rumores. Atividade designada a uma pessoa específica que fazia tudo manualmente	Eram abordadas algumas páginas específicas do Facebook e alguns sites para construção do CLIPPING semanal que reunia notícias para divulgação interna. Tudo isso de maneira manual	Busca em diversas fontes de informações e agregadas através de algumas ferramentas. Elaboração do clipping para divulgação interna
Como as Mídias Sociais são utilizadas na busca por rumores?	Superficialmente através da utilização de alguns agregadores de notícias	Superficialmente utilizada e de maneira manual.	Superficialmente e sem foco em conteúdo específico
Com que frequência elas são utilizadas?	O uso direto a Mídias Sociais é deixado de lado. Há apenas análise de informações trazidas por algumas ferramentas.	Esporádico. Uso ainda está em desenvolvimento, exploratório e superficial	Raramente. Uso em alguns eventos de massa
Quais as dificuldades encontradas?	Mídias Sociais são difíceis de trabalhar, há muita informação útil, mas também muita informação inútil que não dá para aproveitar. Perda de informações durante monitoramento em tempo real, sem poder recuperar seu conteúdo.	Dificuldade na análise manual de um grande volume de informações	Acompanhamento manual das informações. Sem parametrização por palavras-chave. Não é usada ferramenta que apoie a atividade

Através desta entrevista foi possível observar o esforço realizado pelas equipes para reunir informações que auxiliem na realização do processo de detecção de rumores em saúde pública. Com o trabalho de poucos profissionais dedicados, esta atividade é realizada através da utilização de alguns poucos recursos computacionais para reunir as diversas informações que serão analisadas.

Mesmo diante do desafio em identificar as informações úteis nas Mídias Sociais, a equipe responsável pela atividade de detecção de rumores em saúde pública se esforça para observar estas informações em tempo real, mesmo que seja de maneira superficial. Por não haver repositórios locais para este grande volume de informações, a perda de conteúdo relevante possui uma alta probabilidade.

Através das respostas obtidas nesta entrevista e da realização da pesquisa bibliográfica, esta pesquisa levantou os requisitos para construção do Framework Conceitual proposto (PRIETO-DÍAZ, 1990).

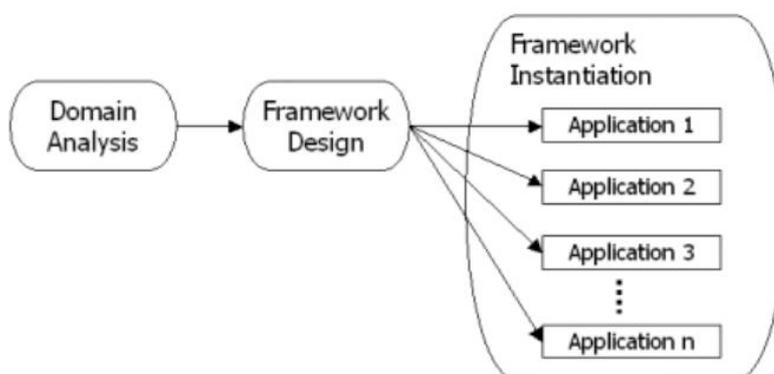
5 FRAMEWORK PARA APOIO À DETECÇÃO DE RUMORES

Através da revisão bibliográfica e entrevista não estruturada, realizadas com os especialistas, identificou-se que as Mídias Sociais têm potencial para o fornecimento de informações relevantes e oportunas sobre saúde, tanto através do texto postado por um usuário quanto pelos seus metadados. Contudo, para realização de contribuições relevantes junto às atividades de vigilância epidemiológica, o grande volume de informações disseminadas por essas Mídias, necessita de um tratamento que seja capaz de solucionar alguns problemas: identificação das informações de maior relevância para as atividades de vigilância para posterior leitura; falta de estrutura das mensagens; uso da linguagem coloquial e símbolos comumente utilizados por seus usuários.

O uso de ferramentas para apoiar uma melhor utilização do potencial existente nessas informações é uma realidade atualmente. Deste modo, busca-se criar ferramentas para melhorar ainda mais o uso das Mídias Sociais como fonte de informações relevantes e oportunas, principalmente as que são relacionadas à saúde. No entanto, para viabilizar a elaboração de sistemas em um domínio complexo, é útil a criação de um framework conceitual que reúna as funções, elementos a serem estudados, suas interações e tudo que contribua para a compreensão do contexto abordado (MILES; HUBERMAN, 1994).

Segundo MARKIEWICZ e LUCENA (2000), o processo de desenvolvimento e utilização de um framework é composto por três atividades macro, conforme exibido na Figura 5.

Figura 5 - Processo de Desenvolvimento de um Framework



Fonte: (MARKIEWICZ; LUCENA, 2000)

As atividades macros são descritas a seguir:

- **Análise de Domínio:** Consiste na identificação e organização do conhecimento sobre o domínio do problema para dar suporte à sua descrição e na solução.
- **Projeto do Framework:** Esta etapa consiste na construção de um framework flexível, com partes estáveis e outras variáveis que aumentem sua extensibilidade (MARKIEWICZ; LUCENA, 2000).
- **Instanciação do Framework:** Deve ter a capacidade de ser instanciado uma ou mais vezes em uma mesma aplicação ou em aplicações diferentes (MATTSSON, 2000).

O framework conceitual, apresentado por esta pesquisa, foi projetado para propor o uso de parâmetros para capturar, tratar as mensagens de Mídias Sociais e identificar as informações relevantes ao contexto da busca por rumores sobre uma determinada síndrome. Através da utilização de palavras-chave relacionadas aos sintomas primários de uma determinada síndrome, a ferramenta permite a captura das informações tanto em tempo real quanto em uma janela de tempo no passado, com datas de início e fim pré-estabelecidas. Durante a captura, o framework propõe o tratamento textual das mensagens, deixando-as mais compreensíveis para as etapas seguintes.

Após o armazenamento das mensagens em um repositório, já com tratamento textual realizado, o framework propõe a realização de sua análise através da utilização de novos parâmetros. Estes parâmetros servem para identificar quais são as mensagens de maior relevância para a investigação, sobre a existência de possíveis rumores. Eles também apontam aquelas mensagens que podem ser descartadas para evitar sobrecarga de informações ou ruídos. Assim, as equipes responsáveis pela detecção de rumores em saúde pública recebem apenas as informações com maior relação com a síndrome cuja existência do rumor está sendo investigada. Através destas mensagens, as equipes podem estabelecer a existência ou não de um rumor, sobre uma síndrome em determinada localidade, apresentar os resultados e emitir alerta a outras áreas da saúde quando necessário. Através da elaboração de relatórios e alertas, as outras equipes responsáveis podem gerar ações de investigações de campos e possíveis respostas a um evento de saúde pública mais rapidamente e de maneira mais eficiente.

Nas seções seguintes deste capítulo, serão apresentadas a concepção, as características do framework e as instâncias elaboradas. Sua estrutura está voltada à utilização de parâmetros definidos pelos analistas especialistas.

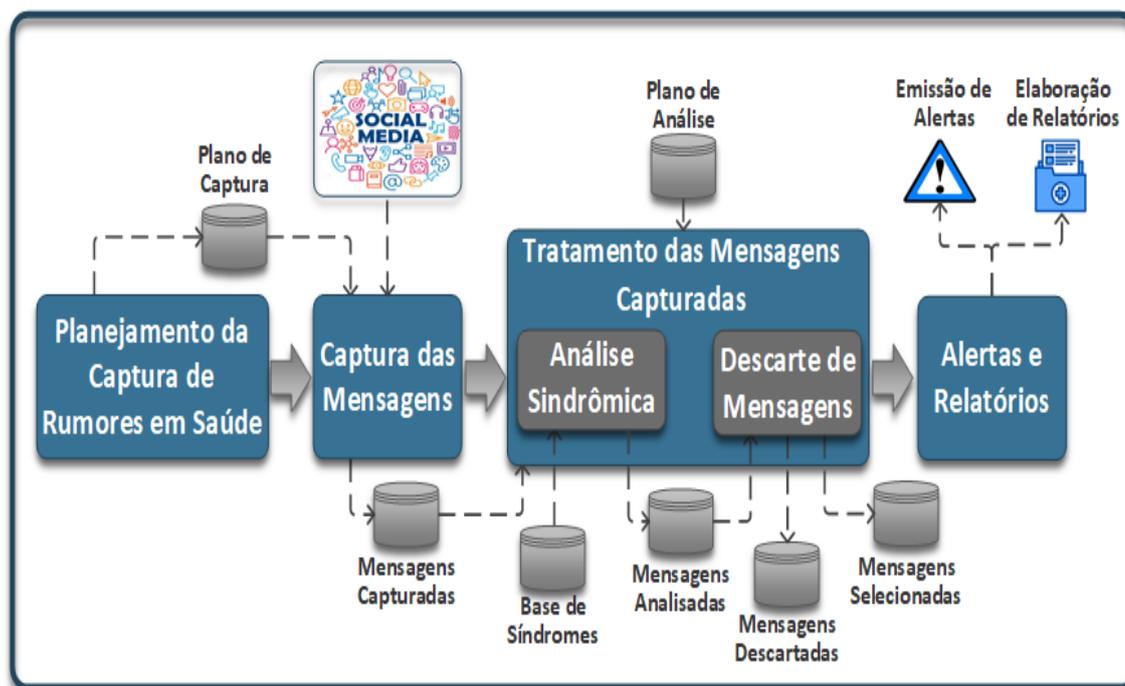
5.1 FRAMEWORK: VISÃO GERAL

O grande volume de informações não estruturadas e muitas vezes com erros, disponíveis nas Mídias Sociais, requerem cada vez mais o uso de recursos da Tecnologia da Informação para melhorar o processo de tratamento e análise destas informações. Para reduzir esse volume a um nível mais interpretável e composto por informações relevantes, tanto a filtragem quanto o descarte de informações irrelevantes são de grande importância (MOONEY; PEJAVER, 2018).

Abordar o imenso volume de mensagens disponíveis, ler, classificar e identificar algo relevante para a vigilância em saúde, são atividades praticamente impossíveis para os profissionais dos CIEVS sem o apoio tecnológico adequado.

Mensagens de grande valor podem ser obtidas através observação de seu conteúdo com parâmetros pré-estabelecidos, tanto para captura quanto para o tratamento. O Framework também propõe atividades relacionadas ao planejamento, tratamento e apresentação dos resultados obtidos durante o processo de busca por um rumor nas Mídias Sociais. A estrutura do framework é apresentada na Figura 6.

Figura 6 - Estrutura do Framework



O framework está distribuído em fases distintas que preservam a flexibilidade durante o desenvolvimento dos sistemas. Isso possibilita que o desenvolvimento das funcionalidades dos sistemas ocorra de maneira evolutiva e preserva os princípios de coesão e acoplamento durante o processo da elaboração da solução final. O artefato também sugere abordagens de trabalho estritamente voltadas aos conteúdos textuais das mensagens. Devido a necessidade de um apoio tecnológico mais complexo, que nem sempre se consegue extrair uma contribuição compensatória para as equipes dos CIEVS em seu processo de busca por rumor sobre uma síndrome, o framework não aborda o tratamento de imagens, vídeos ou áudios.

No intuito de aproveitar ao máximo as contribuições que as Mídias Sociais possam oferecer sobre informações úteis e capazes de direcionar corretamente a investigação sobre a existência de um rumor sobre uma síndrome, este framework apresenta algumas contribuições peculiares, que permitem que as mensagens sejam captadas e trabalhadas o suficiente para alcançar um resultado de nível satisfatório estabelecido pelas equipes dos CIEVS:

- **Manutenção do histórico de informações:** Deve ser mantido em uma base de dados, informações sobre diversos processos de busca por rumor, independente de terem

sido utilizadas concorrentemente ou em momentos diferentes. A existência dessas informações armazenadas, permite a reutilização das informações estabelecidas em cada uma das buscas registradas. Esse reuso permite a reaplicação de um processo de busca de maneira íntegra ou parcial quando houver a tentativa de utilizar partes dessas informações como base para realização de uma nova busca.

- **Automatização de funções:** Em meio a um processo de detecção de rumor nas Mídias Sociais, é necessário um grande esforço para capturar e tratar um volume de mensagens suficiente para estabelecer a existência ou não do rumor procurado. Deste modo, o framework propõe automatizações de determinadas funções realizadas pelas equipes do CIEVS. Ao estabelecerem cada um dos respectivos planos a serem aplicados na fase de captura e tratamento das mensagens, a solução deve aplicar os parâmetros pré-estabelecidos de forma automática. Em consequência disso, essa automatização estabelece classificações das mensagens e descarta aquelas com pouca ou nenhuma contribuição para o processo em andamento. Isso deixa, por fim, aquelas com maior valor e relevância para análise das equipes.

5.2 FUNÇÕES DO FRAMEWORK

Cada uma das fases do framework exibidas na Figura 6 corresponde a uma ou mais funções. A seguir seguem maiores detalhes sobre cada uma delas:

5.2.1 PLANEJAMENTO DA CAPTURA DE RUMORES EM SAÚDE

O framework ressalta a necessidade de um planejamento para o estabelecimento de parâmetros que irão balizar as buscas por informações junto às Mídias Sociais. Os parâmetros definidos devem ser capazes de compreender as características das mensagens consideradas úteis ao andamento da investigação.

Sob a denominação de plano de captura, os parâmetros a serem considerados são:

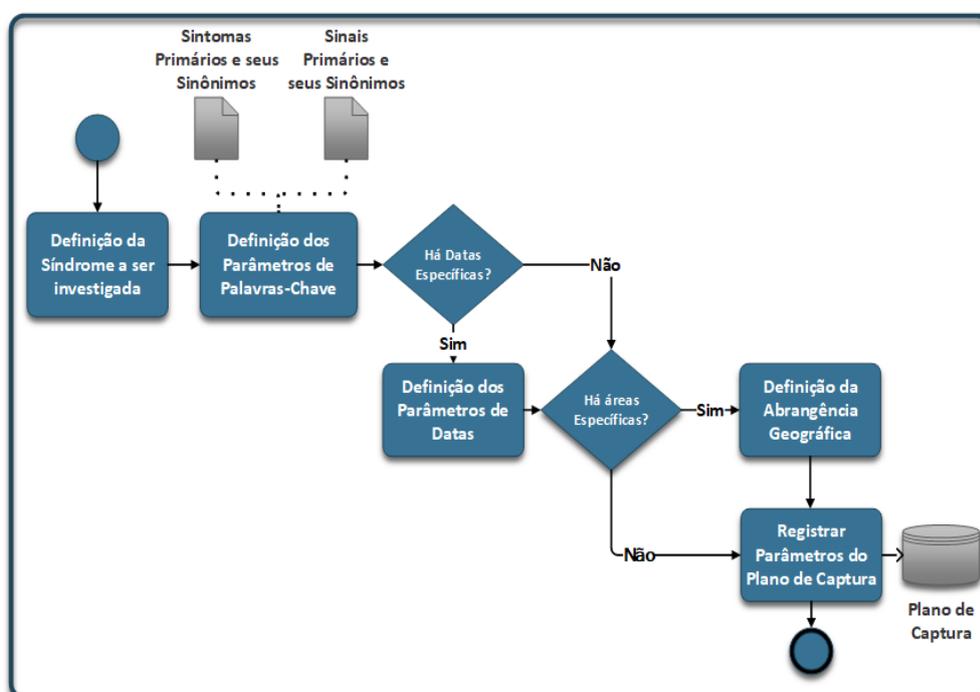
- **Palavras-Chave** – Esse parâmetro contempla a aplicação de palavras-chaves que abordem os sintomas e sinais considerados primários ou obrigatórios de uma

síndrome com seus respectivos sinônimos e variações regionais. Desta maneira, melhora a capacidade de identificar informações importantes, mesmo quando descritas de maneira diferente, conforme a região abordada (regionalização das palavras);

- Posicionamento Geográfico – Esse parâmetro permite a delimitação de uma região geográfica para captura das mensagens, através da identificação de um ponto geográfico e o raio de abrangência.
- Período de Abrangência da Captura – A data de postagem da mensagem é importante para a investigação. Portanto, através deste parâmetro, é possível a captura de mensagens postadas em um período específico.

A Figura 7 exibe o fluxo de estabelecimento desses parâmetros para compor o plano de captura.

Figura 7 - Fluxograma para elaboração do Plano de Captura



5.2.2 CAPTURA DAS MENSAGENS

O framework contempla a captura de mensagens junto às Mídias Sociais de duas maneiras: captura em tempo real ou captura de mensagens já postadas. No primeiro caso, a captura ocorre conforme o surgimento de mensagens, que estejam em conformidade com os parâmetros estabelecidos. No segundo, a captura foca em mensagens já postadas, através do estabelecimento de um período passado, permitindo a análise histórica do comportamento das mensagens.

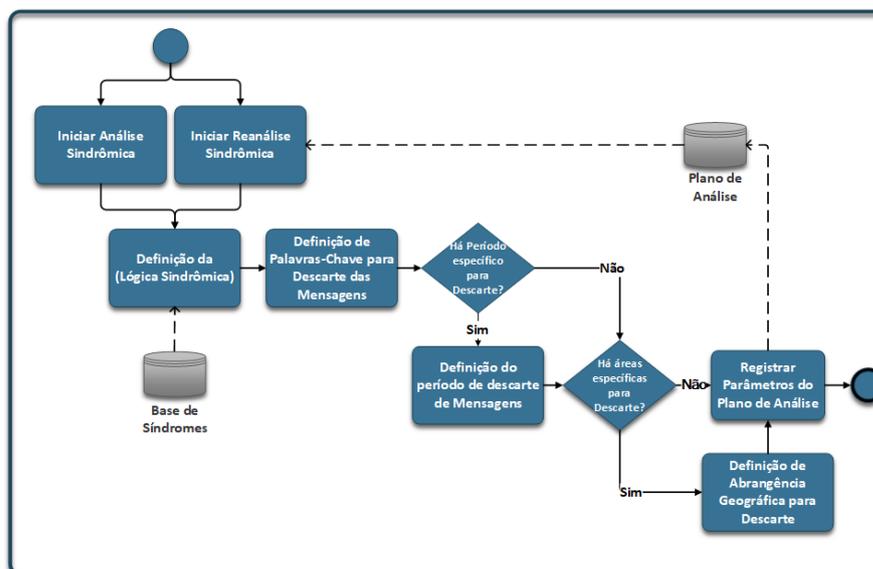
Independente do caso, a solução desenvolvida com base no framework deve ser capaz de realizar uma limpeza no conteúdo textual capturado nas Mensagens de Mídias Sociais antes do seu armazenamento no repositório apropriado para facilitar a compreensão do seu conteúdo. A organização sugerida refere-se a: retirar caracteres característicos de múltiplas linhas (“line feeds”, “carriage return”) e marcadores de tabulação; padronizar todo conteúdo em letras maiúsculas ou minúsculas; retirada de acentos e caracteres especiais; retirada dos links para mídias e hashtags no conteúdo.

5.2.3 TRATAMENTO DAS MENSAGENS

O objetivo desta etapa é analisar cada uma das mensagens capturadas, classificando-as pela observação de seu relacionamento com a síndrome.

Neste momento, o framework também observa a necessidade de um planejamento para estabelecer um novo plano, mas desta vez este plano tem o objetivo de analisar as mensagens capturadas anteriormente. Sob a denominação de plano de análise, o agrupamento dos parâmetros estabelecidos para o tratamento das mensagens se divide em duas outras funções distintas, como é exibido na Figura 8. Neste caso, a solução desenvolvida deve ser capaz disponibilizar tanto a função de tratamento das mensagens como as duas outras que ela contém internamente.

Figura 8 - Fluxograma para elaboração do Plano de Análise



5.2.3.1 ANÁLISE SINDRÔMICA

Através desta função, a solução, baseada no framework, classifica as mensagens segundo sua relação com a síndrome que está sob foco da investigação. Para possibilitar a aplicação desta função, o framework considera que a solução desenvolvida deve acessar um repositório de dados pré-existente e construído por especialistas na área da saúde. Denominado como base de síndromes, este repositório contém a relação de cada uma das síndromes mapeadas pela área da saúde com seus sintomas, sinais e respectivos sinônimos e variações regionais considerados secundários ou facultativos que estão dispostos de maneira logicamente organizada.

O framework considera que cada organização lógica distinta das informações secundárias sobre uma síndrome, representam diferentes lógicas sindrômicas que podem ser utilizadas para identificar informações relevantes durante uma análise. Através da utilização dos operadores lógicos AND, OR e NOT é possível estabelecer uma ou mais lógicas sindrômicas sobre uma mesma síndrome, bastando apenas alterar a disposição das informações secundárias desta síndrome e dos operadores. Ao considerar esse rearranjo do conteúdo das lógicas sindrômicas registradas, o framework contempla possíveis atualizações de parâmetros e diferentes níveis de restrição lógica sobre a análise.

A Tabela 2 exemplifica duas lógicas sindrômicas voltadas a identificação de mensagens relacionadas com a síndrome diarreica e respiratória, respectivamente.

Tabela 2 - Exemplos de Lógicas Sindrômicas

Nomenclatura da Síndrome	Sintoma / Sinal Secundário OR Sinônimo do Sintoma / Sinal Secundário	Op. Lógico	Sintoma / Sinal Secundário OR Sinônimo do Sintoma / Sinal Secundário	Op. Lógico	Sintoma / Sinal Secundário OR Sinônimo do Sintoma / Sinal Secundário	Op. Lógico	Sintoma / Sinal Secundário OR Sinônimo do Sintoma / Sinal Secundário	Op. Lógico	Sintoma / Sinal Secundário OR Sinônimo do Sintoma / Sinal Secundário
Diarreica	(febre OR hipertermia)	AND	(febre OR quentura)	AND	(dor abdominal OR dor de barriga)	AND	(dor abdominal OR colica)	AND	(nausea OR enjojo)
Respiratória	(mal estar)	AND	(cefaleia OR cefalia)	AND	(cefaleia OR dor de cabeça)	AND	(dor no corpo OR algia)	AND	(dor de garganta OR garganta arranhada"

Através da base de síndromes, o framework registra uma ou mais lógicas sindrômicas de uma mesma síndrome. A cada análise das mensagens capturadas, ele seleciona uma das lógicas para ser utilizada como critério da análise em andamento e irá compor o Plano de Análise que está sendo aplicado. Desta maneira, o framework considera que a combinação estabelecida por especialistas na área da saúde é capaz de identificar as mensagens de maior relação com determinada síndrome investigada e que trarão maior valor à investigação realizada.

As mensagens cuja relação com a síndrome foi identificada serão agrupadas sob a classificação de Mensagens Analisadas.

5.2.3.2 DESCARTE DE MENSAGENS

O framework estabelece que a solução desenvolvida deve ser capaz de identificar as mensagens cuja contribuição para o processo de investigação da existência de um rumor é baixa ou inexistente.

Para identificar as mensagens descartáveis e retirar possíveis ruídos contidos nas informações que serão utilizadas pelos especialistas em saúde, esta parte do framework estabelece previamente novos parâmetros que serão aplicados sobre as mensagens. Com a utilização destes parâmetros durante o tratamento das mensagens, o Framework reduz a ocorrência de possíveis descartes de informações relevantes durante uma investigação sobre a existência de um rumor.

Posicionados logo após o processo de análise sindrômica, o descarte de Mensagens irá atuar sobre as mensagens analisadas, ou seja, mensagens cuja relevância já foi identificada anteriormente. Desta maneira, ao aplicar seus parâmetros sobre essas mensagens, o Framework consegue realizar um refinamento das mensagens.

Juntamente com a lógica sindrômica, os parâmetros de descarte irão compor o Plano de Análise utilizado para a etapa de tratamento de mensagens. Os parâmetros de descarte devem ser capazes de identificar informações que tenham alguma característica que não interessa para a investigação. A solução deve permitir a classificação correspondente através do uso dos seguintes parâmetros:

- **Palavras-Chave** – Esse parâmetro contempla a aplicação de palavras-chaves que possam identificar mensagens já analisadas que sejam consideradas irrelevantes para o processo. Palavras como, “jogo” ou “genocida” podem identificar Mensagens Analisadas que estejam relacionadas a um jogo de algum esporte que esteja em andamento ou a uma crítica política quanto a um governo, conforme exemplifica a Tabela 3.

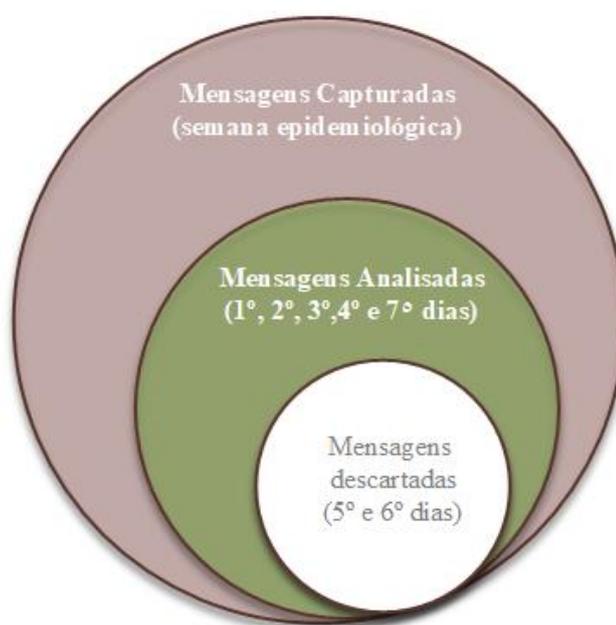
Tabela 3 - Exemplo de Tweets Descartados

Palavra Chave de Descarte	Tweet descartado	Motivo do Descarte
"Jogo"	"mano ultimamente estava cagando se o vasco ganhava ou perdia mas o jogo de hoje pqp um crime barbaro em sao januario hoje"	Posicionamento do remetente com relação a uma partida de futebol
"genocida"	"ssarahandrade apoia nada seguidora e fa d genocida do bolsonaro espero que vc nunca sinta falta de ar como tivesse se afogando e sentido a morte perto tu devia sumir da midia voce desde o bbb e um desservico a sociedade voce ate assumiu que frequentava varias festas clandestinas tu e nefasta "	Posicionamento do remetente com relação a um cenário político

- **Remetente não confiável** – Este parâmetro corresponde à identificação de um autor cujas mensagens são consideradas desprezíveis para o processo. Através deste parâmetro, a solução permitirá que o sistema possa descartar as mensagens de qualquer remetente identificado como não confiável por outras fontes de informações ou pela observação do seu histórico de mensagens que geralmente não contribuem ou prejudicam a investigação que esteja em andamento. Desde modo, é possível o descarte das mensagens cujo autor é conhecido por disseminação de “fake News” ou por ser um “bot” gerador de inúmeras mensagens consideradas irrelevantes para a investigação.

Período de Descarte – Este parâmetro permite o descarte de mensagens postadas dentro de um período estabelecido. Desta maneira, é possível o descarte de mensagens provenientes de datas específicas que não sejam interessantes para a análise em andamento, ou seja, o Framework permite que as avaliações dos especialistas sejam realizadas apenas nas mensagens relevantes compreendidas dentro do período desejado, reduzindo assim, o trabalho de observação a ser realizado. Por exemplo, ao realizar o tratamento das mensagens capturadas durante uma semana epidemiológica, foi estabelecido que as mensagens do 5º e 6º são irrelevantes para a análise em andamento. Então, após a identificação das mensagens analisadas, aplica-se o parâmetro de período de descarte para o 5º e 6º dia, deixando apenas as mensagens dos demais dias disponíveis para os especialistas no restante do processo. A Figura 9, mostra como esses conjuntos de mensagens interagem entre si.

Figura 9 - Descarte de Mensagens por Período



- Posicionamento Geográfico – Este parâmetro permite o descarte de mensagens que tenham sido postadas em determinada região geográfica estabelecida. Semelhante ao período de descarte, este parâmetro também flexibiliza o descarte de mensagens segundo sua localização geográfica. Neste momento, o Framework recebe nova coordenada geográfica dentro da área estabelecida pelo plano de captura e um raio (Km) para identificação da área que se deseja desconsiderar as mensagens. A Figura

10, exemplifica uma investigação de mensagens originadas em uma área da Zona Sul do Rio de Janeiro que desconsidera as mensagens analisadas originadas no bairro da Gávea.

Figura 10 - Descarte por Posicionamento Geográfico



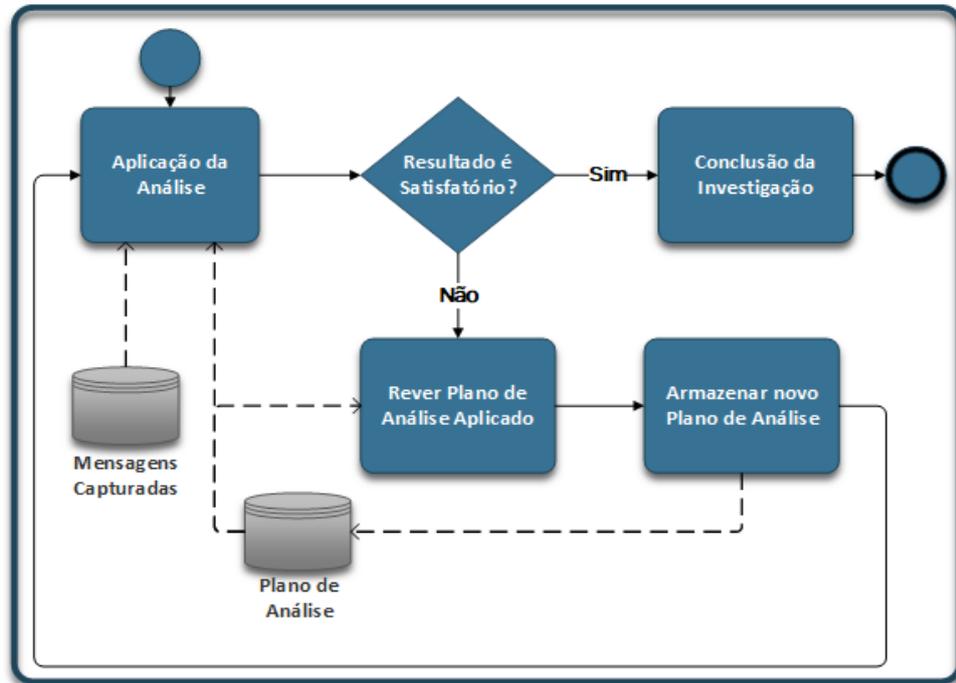
5.2.3.3 REANÁLISE DAS MENSAGENS

Ao final do tratamento das mensagens, o resultado obtido pode não ser satisfatório para as equipes responsáveis tirarem suas conclusões e diagnosticarem a existência ou não do rumor investigado. No intuito de auxiliar na resolução deste problema, a solução desenvolvida deve possibilitar a revisão e refinamento dos parâmetros contidos no plano de análise para realizar uma nova análise sobre essas mensagens.

Denominado como reanálise das mensagens, esta parte do Framework permite que sobre um mesmo volume de mensagens capturadas por um plano de captura sejam aplicados diferentes planos de análise. Isto possibilita que a solução explore, sob diferentes parametrizações e obtenha diferentes resultados de análises.

O fluxo apresentado na Figura 11 refere-se a um exemplo da realização do refinamento dos parâmetros estabelecidos pela lógica sindrômica.

Figura 11 - Fluxo de Reanálise das Mensagens



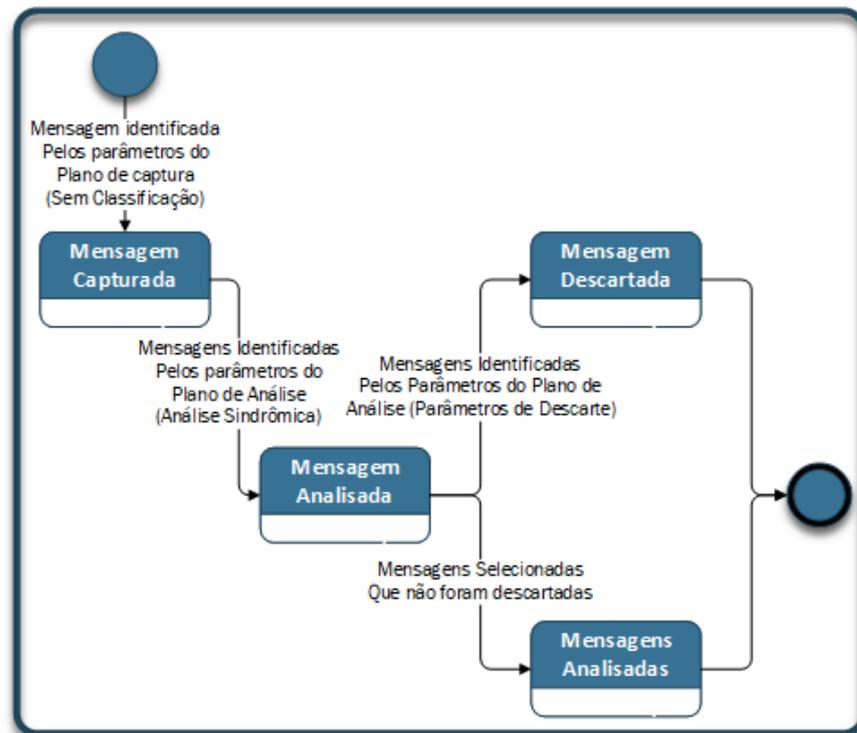
5.2.4 ALERTAS E RELATÓRIOS

Para a disseminação das informações inerentes ao processo, o framework estabelece que a solução deva possuir as seguintes funções: Elaboração de relatórios e emissão de Alertas.

Através da construção e disponibilização de gráficos e relatórios, a solução irá disseminar o conteúdo atualizado e o seu resultado. Essas informações podem ser organizadas em relatório próprio da solução ou pode ser disponibilizada como insumo para outros modelos de relatórios já existentes.

A Figura 12 exemplifica um relatório que apresenta os resultados adquiridos com a execução de um plano de captura. Neste relatório, é possível observar o resultado de um plano captura aplicado, ler o conteúdo das mensagens capturadas e através de um gráfico de palavras que mostra um resumo visual (CHANDRAPAUL *et al.*, 2019) do volume de mensagens que cada palavra-chave utilizada em um plano identificou junto a Mídia Social.

Figura 13 - Classificações das Mensagens



As classificações correspondentes às fases do framework serão descritas a seguir:

- **Mensagens Capturadas:** Estas mensagens são mensagens resultantes da aplicação dos parâmetros descritos em um plano de captura junto a API das Mídias Sociais. Em concordância com esses parâmetros, estas mensagens são trazidas e armazenadas em um repositório.
- **Mensagens Analisadas:** Essas mensagens correspondem às Mensagens Capturadas, cuja relação com a síndrome foi identificada segundo parâmetros estabelecidos em um plano de análise. Após aplicação dos parâmetros combinados que compõem a lógica síndrômica adotada, as mensagens capturadas recebem essa nova classificação.
- **Mensagens descartadas:** Após identificação das Mensagens Analisadas, o Framework observa novamente estas mensagens para identificar aquelas com pouca ou nenhuma contribuição para o processo de detecção de rumores. Após a aplicação dos parâmetros de descarte sobre as Mensagens Analisadas, aquelas que tiveram relação com tais parâmetros, recebem a classificação de Mensagens Descartadas.

Estas mensagens não são deletadas do repositório em si, mas são classificadas como descartadas para o processo em andamento. Isso permite que sejam revisitadas em outro momento através de uma reanálise.

- Mensagens Seleccionadas: Esta classificação é dada as Mensagens Analisadas que não foram descartadas. Esta classificação é dada as Mensagens consideradas relevantes para a investigação segundo a lógica sindrômica. Essa classificação agrupa as mensagens de maior valor para o processo de investigação sobre a existência de rumores sobre uma síndrome, sendo disponibilizada às equipes de vigilância em saúde para estabelecimento do resultado.

5.2.6 REPOSITÓRIOS DO FRAMEWORK

O framework estabelece que a solução construída deve manter alguns repositórios para manipular as mensagens capturadas e os parâmetros utilizados. O armazenamento em repositórios possibilitará que o sistema mantenha o histórico dessas informações e sua reutilização. A seguir seguem as descrições dos repositórios:

- Plano de Captura – Este repositório armazena os parâmetros estabelecidos durante a fase de planejamento da captura em rumores descrita pelo framework. Esses parâmetros definem critérios de captura para identificar informações úteis junto às Mídias Sociais. Composta por palavras-chaves, informações de geolocalização e período de abrangência da captura. Através deste repositório é possível a reutilização do plano na íntegra e obtenção de novos dados caso necessário.
- Mensagens Capturadas – Através deste repositório serão armazenadas as mensagens capturadas e relacionadas aos parâmetros estabelecidos em um plano de captura aplicado.
- Plano de Análise – Este repositório armazena as relações entre os parâmetros armazenados em dois outros repositórios: Base de Síndromes e Critérios de Descarte. Os conteúdos destes dois repositórios irão compor os parâmetros do Plano de Análise descrito pelo Framework.

- Base de Síndromes – Este repositório armazena os sintomas, sinais e respectivos sinônimos e variações regionais consideradas secundárias de cada uma das síndromes mapeadas pela área da saúde. Dispostos de maneira logicamente organizada, estas informações compõem a lógica sindrômica que será aplicada às mensagens durante o tratamento. Cada uma dessas lógicas está relacionada a uma síndrome específica e devido a existência de diversas variações lógicas possíveis, uma mesma síndrome pode ter diversas lógicas sindrômicas associadas.
- Descarte de Mensagens – Este repositório armazena os parâmetros que serão aplicados às mensagens durante seu tratamento e após sua análise sindrômica. Esses parâmetros definem os critérios de descarte das mensagens durante o processo descrito pelo Framework. Composta pelos parâmetros de palavras-chave, remetente das mensagens, data das mensagens e geolocalização.

6 AVALIAÇÃO CONCEITUAL DO FRAMEWORK

Com objetivo de entender se o framework proposto nesta pesquisa é capaz de trazer melhorias significativas, ele foi apresentado aos profissionais e especialistas em saúde que apoiaram esta pesquisa. Após a apresentação e discussão de cada uma das partes sugeridas pelo framework, os especialistas responderam um questionamento sobre a proposta.

Durante todo o processo de apresentação, discussão e questionamentos, foi possível perceber uma boa aceitação da proposta elaborada. As discussões apontaram para uma escolha assertiva nos parâmetros estabelecidos para o Framework proposto que se mostraram úteis ao processo atual. O armazenamento das mensagens capturadas e o processo de análise proposto pelo framework, foi ressaltado como positivo, pois permite que sejam analisadas um maior volume de mensagens com emprego de menor esforço pelas equipes e em momento oportuno.

A proposta de utilizar as palavras-chaves relacionadas aos sintomas da síndrome investigada foi considerada um ótimo direcionador para identificar e analisar mensagens durante o processo de detecção de rumores. Os parâmetros sugeridos pelo Framework foram considerados úteis, mas através desta entrevista também foi identificado alguns pontos de melhorias que foram: a inserção do parâmetro de posicionamento geográfico e período como critérios de descarte. Segundo os especialistas, a possibilidade de excluir mensagens de uma subárea geográfica ou de algum período irrelevante é interessante para a investigação, pois aproxima ainda mais as funcionalidades do Framework de uma situação cotidiana das equipes de saúde.

A avaliação do Framework pelos entrevistados foi positiva, sendo considerada uma grande contribuição para redução do esforço atual despendido para análise das mensagens de Mídias Sociais para identificação de rumores em saúde pública.

Nesta fase de concepção, a equipe de especialistas foi composta por 3 profissionais e a resposta dos questionamentos será exibida na Tabela 4.

Tabela 4 - Compilação da Avaliação conceitual do Framework

QUESTÕES	PROFISSIONAL 1	PROFISSIONAL 2	PROFISSIONAL 3
Tempo experiência (Anos)	3 anos	1 ano	9 anos
O estabelecimento de parâmetros como: palavras-chave, localização e data da mensagem contribuem para melhoria dos resultados no processo de detecção de rumores?	São ótimos parâmetros. Contemplar regionalismo e poder formar algo como um Dicionário de Palavras Chaves com os parâmetros utilizados é muito interessante.	É uma grande contribuição para o processo de detecção. Poder armazenar os parâmetros em planos é uma ótima organização que possibilita uma consulta posterior.	Ótimas categorias. O tempo, pessoa e lugar são sempre um norte seguido nas investigações
O armazenamento de grandes volumes de informações em uma base de dados, contribui para realização de novas buscas?	Tomando os devidos cuidados com os dados considerados sensíveis, representa uma grande contribuição.	Um grande volume de informações é interessante se utilizado a automação proposta.	É uma boa estratégia, mas deve ter cuidado com as especificidades de cada síndrome e seus sintomas
Como o estabelecimento de sintomas, considerados primários, de uma síndrome como parâmetro pode contribuir para a captura das mensagens?	Bem prático e traz maior facilidade pela realização de uma busca mais específica	É bem interessante, desde que esteja bem amarrado ao escopo da síndrome investigada	É legal. Havendo um processo automático que suporte a manipulação dessas informações e dois refinamentos necessários, isso será ótimo.
De que maneira a utilização de sintomas, considerados secundários, como parâmetros pode ajudar a correlacionar as mensagens com a síndrome cujo rumor está sendo buscado?	Poder trazer as mensagens e depois fazer uma análise mais específica ajuda e muito. Realizar uma análise mais profunda, delimita e direciona melhor o que você quer.	Poder realizar diferentes análises de síndromes sobre as mensagens capturadas	É uma boa estratégia. A implantação dessa correlação nos serviços é de grande ajuda. Analisar síndromes e não doenças é muito bom
De que maneira o estabelecimento de parâmetros para descarte de mensagens já analisadas pode ajudar na filtragem daquelas mais relevantes?	Os parâmetros propostos são extremamente relevantes. Poder observar os resultados e fazer refinamentos será bastante útil e será utilizado o tempo inteiro	Ajuda no manejo das informações e a efetuar melhores descarte de informações ruins	Poder aprimorar os critérios de descarte irá resultar em informações melhores e mais específicas sobre o que se deseja. Isso ajuda muito

De que maneira a estrutura geral apresentada pelo framework pode contribuir para o processo de detecção de rumores de saúde em Mídias Sociais?	Grande contribuição. Não só a nível nacional, mas também a nível internacional com compartilhamento de tecnologia	Indícios apontados pela solução irão dar subsídios aos técnicos para apontar a existência ou não de rumores.	Contribui muito pois agiliza o processo de identificação de rumores
O que mais lhe agradou na proposta?	A automação do processo de averiguação de posts é muito boa.	Refinamento dos parâmetros, possibilidade dos descartes das mensagens irrelevantes, emissão de alertas.	A reutilização e refinamento de parâmetros é bem interessante. Aplicando em localidades diferentes, pode ajudar a acompanhar algumas dinâmicas epidemiológicas
Sugeriria alguma mudança?	Sem sugestões	Acréscimo de parâmetro de localização também para os parâmetros de descarte	Sem Sugestões

6.1 INSTANCIACÃO DO FRAMEWORK

Após o desenvolvimento e ajustes finais do framework, esta pesquisa realizou uma avaliação sobre a concepção da solução sugerida. O objetivo é avaliar se o framework, através de dados reais, apresenta resultados relevantes que sejam capazes de contribuir com as equipes dos CIEVS.

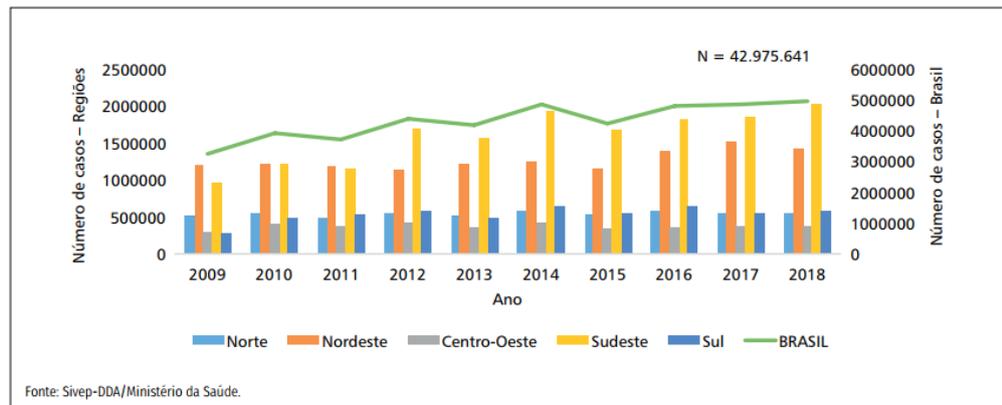
O método de avaliação estabelecido foi a elaboração de duas instâncias do processo de identificação de rumor junto às Mídias Sociais com uso de dados reais e recentes. Com intuito de conseguir resultados mais relevantes sobre síndromes bastante investigadas e capazes de contribuir para a avaliação do Framework, esta pesquisa estabeleceu as instâncias de investigações sobre rumores das seguintes síndromes:

- Síndrome de Diarreia:

Considerada um problema de saúde global, esta síndrome que registra diversos casos (BÜHLER; BONILHA; OLIVEIRA JUNIOR, 2021). Geralmente autolimitada, acompanhada por vômitos, febre, esta síndrome pode levar a mortes por desidratação (BRASIL *et al.*, 2010). Esta síndrome possui muitas ocorrências em países em desenvolvimento e se proliferam sobretudo em cidades com carência de saneamento básico (PEREIRA; MACHADO, 2014; SOUZA *et al.*, 2012).

Segundo o Ministério da Saúde e a Secretaria de Vigilância em Saúde (2021), no Brasil, entre 2009 e 2018 houve uma média de 4 milhões de casos em mais de 33 mil unidades sentinelas, sendo 36,99% só na região Sudeste conforme exibido no Gráfico 2.

Gráfico 2 - Distribuição dos casos notificados em unidades sentinelas por região e ano, Brasil 2009 a 2018



Fonte: (MINISTÉRIO DA SAÚDE; SECRETARIA DE VIGILÂNCIA EM SAÚDE, 2021)

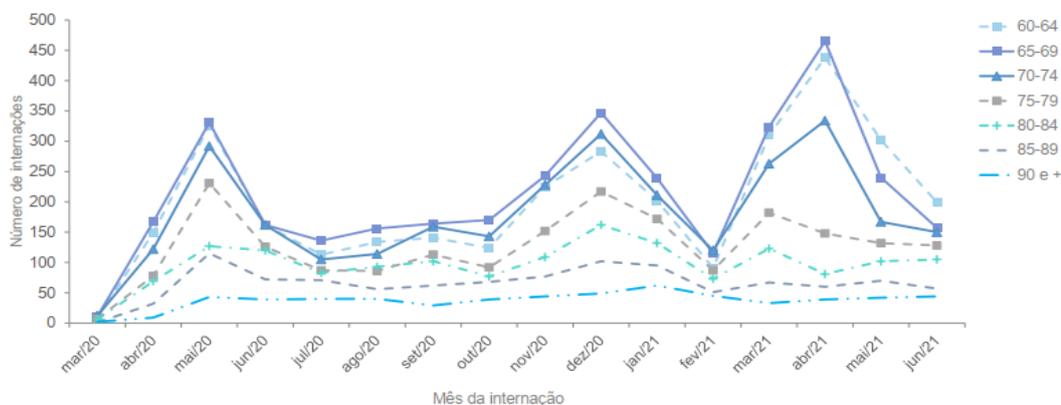
- Síndrome Respiratória:

Considerada uma síndrome infecciosa que vem sendo monitorada de perto desde 2009 no Brasil (BASTOS *et al.*, 2020; PEIRIS *et al.*, 2003). Segundo o Boletim do Sistema InfoGripe, foi observado a interrupção de uma tendência de queda para os casos da síndrome em diversas localidades, entre elas o Rio de Janeiro (FIOCRUZ, 2021). Nos últimos meses, o Estado mostrou picos no volume de casos que superam registros durante o ano de 2020 (BRASIL; MINISTÉRIO DA SAÚDE; SECRETARIA MUNICIPAL DE SAÚDE, 2021), conforme mostra o Gráfico 3.

Gráfico 3 - Casos de SRAG por faixa etária

CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE POR FAIXA ETÁRIA.

Internações em unidades SUS a partir de 60 anos, 2020-2021



Fonte: Fonte: SUBGERAL / Complexo Regulador / Plataforma SMSRIO
Atualizado em: 15/07/2021

Fonte: (BRASIL; MINISTÉRIO DA SAÚDE; SECRETARIA MUNICIPAL DE SAÚDE, 2021)

Após observação desses dois cenários distintos sob duas síndromes diferentes, porém relevantes devido aos seus impactos no sistema de vigilância, foram elaboradas instâncias de investigação sobre rumores a respeito da Síndrome Diarreica e Síndrome Respiratória. Com isso, busca-se mostrar a variabilidade dos resultados e a utilização de informações conforme proposto pelo framework.

6.1.1 ESTRUTURA DAS INSTÂNCIAS

As instâncias que serão apresentadas a seguir, foram projetadas de maneira que pudessem mostrar a capacidade das funcionalidades do framework ao lidar com situações do mundo real e com informações reais, representando um passo importante de uma pesquisa (HEVNER ALAN, 2007).

A Mídia Social escolhida para realização da captura de informações que compuseram as instâncias propostas foi o Twitter. Esta escolha foi feita devido a sua grande quantidade de informações relacionadas à saúde, publicadas pelos seus mais de 500 milhões de usuários pelo mundo; sua disponibilidade de informações em tempo real; suas informações relacionadas a data de publicação das mensagens e a localização

geográfica de sua publicação, características relevantes para composição das instâncias elaboradas.

Em seguida, foi estabelecido o período de 7 dias consecutivos como abrangência da captura das mensagens. Segundo os especialistas, este período é considerado razoável para observação do comportamento epidemiológico de uma determinada região. Outra indicação para o estabelecimento deste período, foi sua utilização pelo Boletim Epidemiológico. Por ser uma publicação de caráter técnico-científico, de acesso livre, em formato eletrônico e com periodicidade semanal para casos de monitoramento de síndromes sazonais, o boletim representa uma ferramenta de disseminação de informações que contribuam para a Saúde Pública do Brasil (MINISTÉRIO DA SAÚDE; SECRETARIA DE VIGILÂNCIA EM SAÚDE, 2022).

Após a definição destas características, foi criada uma conta de desenvolvedor gratuita no Twitter. Esta conta permite acesso a um histórico de 7 dias consecutivos de mensagens através de sua API na versão 1.1 (TWITTER, 2022). Em seguida, através de um protótipo elaborado na linguagem PHP com utilização do banco de dados MySQL, este protótipo realizou a captura e armazenamento das mensagens segundo os parâmetros estabelecidos. A execução do protótipo foi através de uma máquina virtual com sistema operacional Windows com 4GB de memória e 50GB de disco.

O contexto aplicado às simulações foi a realização de duas investigações sobre a possível existência de dois rumores distintos em uma mesma área geográfica em um mesmo período. Cada uma dessas investigações aborda investigações sobre existência de rumores sobre síndromes distintas: Síndrome diarreica e Síndrome Respiratória.

Para a construção das instâncias apresentadas por esta pesquisa, foram estabelecidas algumas considerações importantes em cada uma das fases do framework. Para melhor compreensão, esses pontos serão esclarecidos a seguir conforme as fases do framework.

6.1.1.1 PLANO DE CAPTURA

Para construção do plano de captura aplicado a cada uma das instâncias, foi estabelecido o seguinte:

- PALAVRAS-CHAVE

Uma síndrome é composta por diversos sinais e sintomas que quando identificados em um paciente, podem caracterizar a sua existência. Porém, para a indicação da existência de uma síndrome, o paciente não precisa ter todos os sinais e sintomas relacionados. Para isso, basta que sejam identificados alguns sinais e sintomas mais comuns e característicos nos pacientes infectados.

Para estabelecer essas palavras-chave, foi realizado uma pesquisa no Google sobre os sintomas, sinais e sinônimos tanto da síndrome diarreica quanto da respiratória. Em seguida, com apoio dos especialistas, estas informações foram divididas em dois grupos: os sintomas primários e sintomas secundários. Como já descrito pelo Framework, no momento da captura, as informações primárias corresponderão aos parâmetros de palavras-chave que pertencerão ao plano de captura aplicado em ambas as instâncias.

Desta maneira, entende-se que mensagens que contenham ocorrências de pelo menos um elemento deste subgrupo, já possui contribuição relevante para a investigação em andamento.

Durante a procura destas palavras-chave nas mensagens foi aplicado o processo de busca padrão do Twitter, ou seja, cada uma dessas palavras-chave pertencentes ao plano de captura foram relacionadas entre si, através da utilização do operador lógico “OR”, conforme mostra a Tabela 5.

Tabela 5 - Organização das palavras de Captura

Nomenclatura da Síndrome	Sintoma / Sinal Primário OR Sinônimo do Sintoma / Sinal Primário	Op. Lógico	Sintoma / Sinal Primário OR Sinônimo do Sintoma / Sinal Primário	Op. Lógico	Sintoma / Sinal Primário OR Sinônimo do Sintoma / Sinal Primário	Op. Lógico	Sintoma / Sinal Primário OR Sinônimo do Sintoma / Sinal Primário	Op. Lógico	Sintoma / Sinal Primário OR Sinônimo do Sintoma / Sinal Primário
Diarreica	Diarreia	OR	cambras	OR	defecado	OR	disenteria	OR	corrução
Respiratória	tosse	OR	expelir	OR	coriza	OR	muco	OR	pigarro

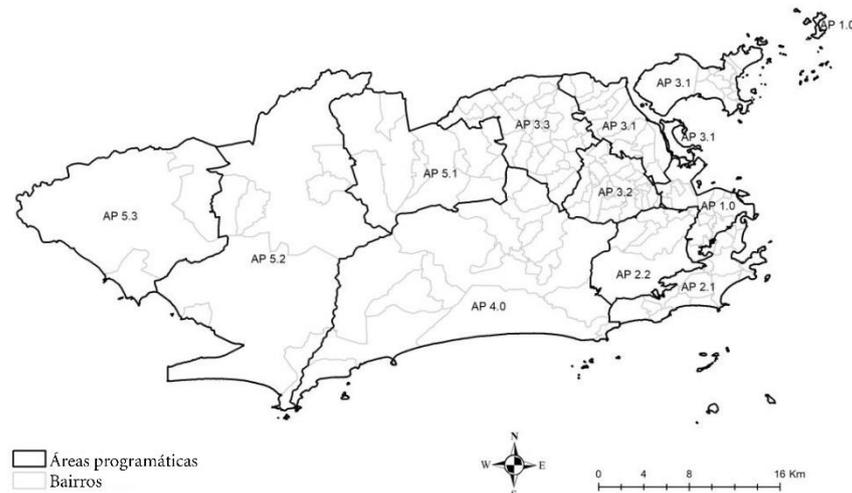
- POSICIONAMENTO GEOGRÁFICO

A cidade do Rio de Janeiro é bastante heterogênea e apresenta diferentes graus de desenvolvimento que causa desigualdades na distribuição da utilização de recursos disponíveis, inclusive nos serviços de saúde. Para melhoria do gerenciamento desses

recursos, em 1993 a Secretaria Municipal de Saúde dividiu geograficamente a cidade em 10 áreas programáticas (APs) (PREFEITURA DA CIDADE DO RIO DE JANEIRO, 2017).

As instâncias que serão apresentadas, irão referenciar uma mesma área geográfica. Para a observação de melhores resultados, a área estabelecida por esta pesquisa foi uma região compreendida na área programática 2.1, que abrange toda zona sul do rio de janeiro (FREITAS, Danielle Amaral De; SOUZA-SANTOS; WAKIMOTO, 2019), segundo exibição feita pela Figura 14.

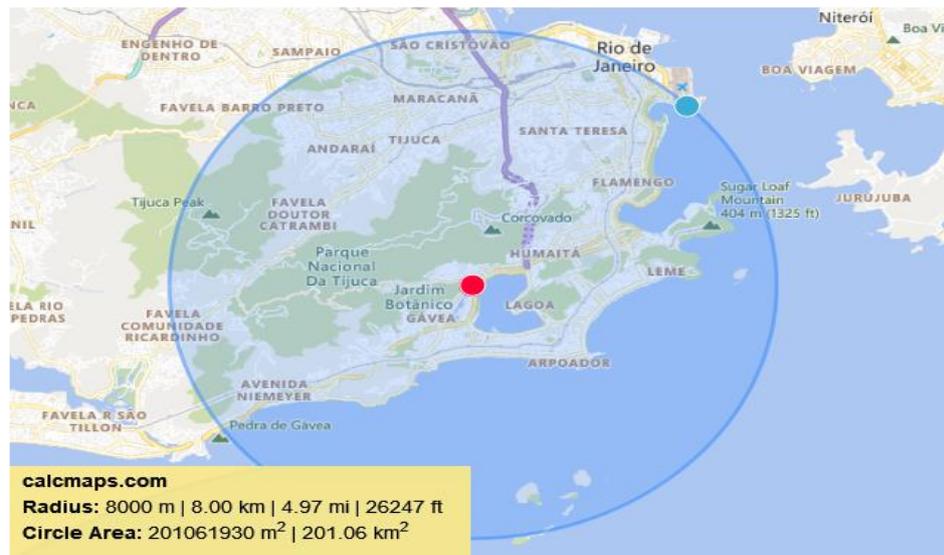
Figura 14 - AP's do Rio de Janeiro



Fonte: (FREITAS, Danielle Amaral De; SOUZA-SANTOS; WAKIMOTO, 2019)

O epicentro desta área foi em meio ao bairro da Lagoa-RJ (latitude: -22.96652147989641, longitude: -43.21803790676029). Estabelecido esse ponto, foi aplicado um raio de 8km de cobertura, estabelecendo uma região circular para captura das mensagens, conforme mostra a Figura 15.

Figura 15 - Plano de Captura - Área de Abrangência



- PERÍODO DE ABRANGÊNCIA DA CAPTURA

Em observação as discussões realizadas com especialistas nesta pesquisa, as instâncias abordadas compreendem um mesmo intervalo de tempo de 7 dias.

Com isso, foi estabelecido a captura das mensagens emitidas dentro deste período específico, ou seja, 7 dias passados a partir da data de aplicação da captura.

Então, no dia 08 de agosto de 2021, foi realizada a captura de mensagens segundo o plano de captura de cada uma das instâncias. Ambos os planos trouxeram as mensagens com emissão compreendidas de 01 a 07 de agosto de 2021. Para uma melhor delimitação das instâncias propostas, e posterior análise das informações capturadas, as instâncias tiveram seu período de abrangência da captura delimitado aos 7 dias anteriores da captura. Desta maneira, as instâncias puderam apresentar a aplicação das funções do Framework sobre um volume de mensagens fixo, sem o constante crescimento observado quando aplicado a captura de mensagens em tempo real.

6.1.1.2 PLANO DE ANÁLISE

Nesta etapa do framework há a utilização dos parâmetros compostos pelo plano de análise: Lógica Sindrômica e Descarte. Eles serão aplicados sobre as mensagens capturadas. Para construção do plano de captura aplicado a cada uma das instâncias, foi estabelecido o seguinte:

- LÓGICA SINDRÔMICA

Resultante da pesquisa no Google realizada, as informações secundárias referem-se a sintomas, sinais e respectivos sinônimos que não são apontados como comuns em pacientes da síndrome segundo os especialistas. Ainda segundo eles, estas informações quando somadas as informações primárias, podem caracterizar ainda mais a presença da síndrome em uma pessoa.

Considerando sua importância na identificação de informações relevantes, neste momento, as informações secundárias serão organizadas logicamente para compor o critério de lógica sindrômica que será aplicada nas duas instâncias apresentadas pela pesquisa.

- DESCARTE DE MENSAGENS

Para estabelecer esses parâmetros foram estabelecidas palavras-chaves ou mesmo radicais dessas palavras para identificação das mensagens que podem ser descartadas no processo.

As palavras-chave de descarte foram estabelecidas através da observação de alguns assuntos gerais e comentados no Twitter (política, futebol etc.). Assim, após a reunião dessas palavras, elas foram aplicadas sobre as mensagens analisadas, identificando aquelas mensagens que serão descartadas no processo.

6.1.2 IDENTIFICAÇÃO DE RUMOR SOBRE SÍNDROME DIARREICA

A primeira simulação apresentada, trata-se de uma investigação para identificar a possível existência de um rumor sobre síndrome diarreica na Zona Sul do estado do Rio de Janeiro.

Essa pesquisa totalizou 49 palavras-chave, conforme exibido na Tabela 6. Desta maneira, a captura de mensagens aumenta sua abrangência para identificar mensagens que possam referenciar os sintomas mesmo que sob outra escrita.

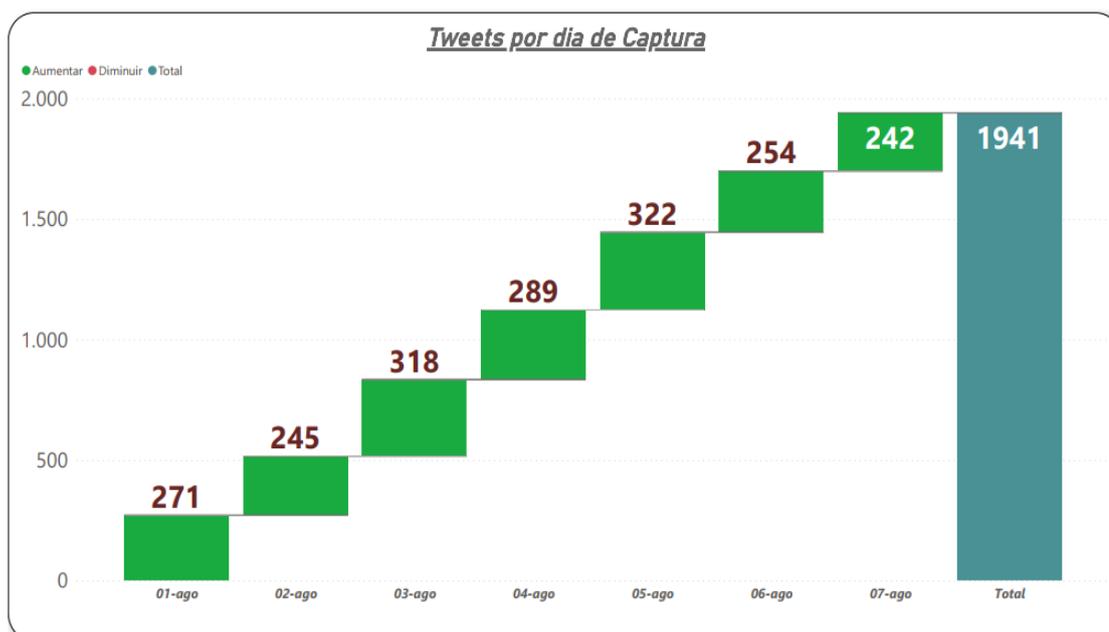
Tabela 6 – Síndrome Diarreica -Plano de Captura-Palavras-Chave

Sinais, Sintomas	Diarreia				
Sinônimos e Variações	afitamento	or	dejeção	or	fezes aquosa
	afito		dejetar		fezes líquida
	borra		desaranjado		fezes mole
	cagado		desaranjo		fezes pastosa
	cagando		destemperado		incontinência
	caganeira		diarreia		ligeira
	cagar		disenteria		maculo
	câmaras		evacua		mal de bicho
	cambras		evacuacao		obrado
	carreirinha		evacuado		obrar
	caseira		evacuando		piriri
	corredeira		evacuar		reira
	corrução		excreção		sedeca
	defecação		excretado		soltura
	defecado		excretar		soltura do ventre
	defecando		excreto		
	defecar		fezes amolecida		

Estabelecidos os parâmetros para o plano de captura, o plano foi aplicado junto a API do Twitter para captura das mensagens correspondentes, compreendidas pelos 7 dias anteriores.

O plano de captura elaborado para essa busca foi aplicado junto a API do Twitter e obteve 1941 tweets capturados, distribuídos pelos dias compreendidos no intervalo de tempo estabelecido conforme exibido no Gráfico 4.

Gráfico 4 – Síndrome Diarreica - Mensagens Capturadas por Dia



Após a captura, todas as mensagens foram submetidas a fase de tratamento para identificar aquelas com maior contribuição para a identificação de um possível rumor sobre síndrome diarreica. A composição dos parâmetros do plano de análise aplicados a esses 1941 tweets será descrita a seguir:

- ANÁLISE SINDRÔMICA

Nesta etapa, foi realizada a análise sindrômica a cada uma das 1941 mensagens capturadas. A lógica sindrômica estabelecida para esta simulação é composta pelos sintomas considerados secundários para síndrome diarreica, seus sinônimos e palavras relacionadas organizados logicamente através do uso de operadores lógicos. Após o levantamento de 59 palavras-chave referentes aos sinais, sintomas e sinônimos secundários, elas foram organizadas logicamente com utilização do operador OR, conforme apresentado na Tabela 7.

Tabela 7 – Síndrome Diarreica -Lógica Sindrômica

Sinais e Sintomas	Febre	OR	dor abdominal	OR	sangue nas fezes	OR	muco nas fezes	OR	vômito	OR	náusea
Sinônimos e Variações	Febril		dor no abdome		fezes escuras		cor das fezes		vomitado		enjoo
	Febrento		dor de barriga		fezes pretas		fezes pretas		vomitando		nojo
	Febrão		dor na barriga		cheiro forte		fezes verdes		vomitando		enojo
	Febrinha		colica		hematoquezia		coco		vomitação		ânsia
	Febrilizar		abdominalgia		melena		merda		golfada		asco
	Febrícula		obstrução		sangramento		bosta		golfo		
	Febriólogo		espasmos		sangrento				bolçado		
	Febroso								jorro		
	Hipertermia	OR		OR		OR		OR	lanço	OR	
	Hiperpirose								êmene		
	hipertermico								desembucho		
	Pirexia								regurgito		
	Quentura								nojo		
	Temperatura corporal Elevada										
	Temperatura corporal alta										

Após a aplicação da lógica sindrômica mostrada na Tabela 7 sobre as 1941 mensagens capturadas, restaram apenas 113 mensagens para continuidade do processo de investigação sobre a existência de rumores sobre síndrome diarreica.

Desta maneira, é possível observar a viabilidade do framework em reduzir o volume de mensagem a ser analisada pelas equipes de especialistas. No princípio, os especialistas teriam que analisar 1941 mensagens, porém, a lógica sugerida nesta instância reduziu esse universo para apenas 113 (6,85% do total). Esse percentual, representa o volume de mensagens com maior relação a síndrome diarreica, cuja existência do rumor está sendo investigada sendo a lógica aplicada.

Consideradas como “Não Classificadas”, as 1808 mensagens (93,15%) não foram descartadas e não foram classificadas. Elas representam os tweets que não se enquadram

nos parâmetros definidos pela lógica sindrômica utilizada. Ter essas mensagens já capturadas e armazenadas em um repositório, permite a realização de uma “reclassificação”, que pode ser realizada a qualquer momento, ou seja, a aplicação de parâmetros mais refinados sobre as mensagens capturadas pode identificar novas mensagens relevantes para investigação.

- **DESCARTE DE MENSAGENS**

Mesmo classificadas como analisadas, as mensagens resultantes do processo de análises sindrômicas ainda passaram por um processo que identifica e descarta as mensagens com menor relevância para a investigação.

Para isto, foram utilizados 57 parâmetros (Tabela 8) para identificação e classificação das mensagens descartadas para esse processo.

Tabela 8 - Parâmetros de Descarte

Parâmetros de Descarte				
as merdas	cagando na cabeça	estar cagado	lula	time
bolsonaro	cagando para	essa bosta que	merda pra cagar	merda que
Borra bosta	cagando pra	falar merda	mesma bosta	militante
pra cagar	cagar pra	faz merda	mesma merda	mito
bosta que	cagando se	ficar cagado	comida caseira	olimpiada
caga regra	cagando solenement e	ficar na merda	defecar pela	os merdas
cagando regra	cagar no pau	fosse merda	é uma merda	passar coco
cagado de algo	comidinha caseira	futebol	se fosse merda	politico
cagando de	eleitor	genocida	stf	presidente
cagando de medo	essa merda	governo	stm	
cagando de rir	esta cagando	ligeira	ta cagado	
cagando e andando	esta merda	ligeira impressão	vasco	

Esses parâmetros foram estabelecidos para identificar as mensagens com pouca ou nenhuma contribuição para a investigação em andamento. Os parâmetros estabelecidos para esta instância tiveram como objetivo o descarte das mensagens que estivessem diretamente relacionadas a outros cenários como política, futebol, entre outros. Com aplicação desses parâmetros sobre as 113 mensagens analisadas, foi possível o descarte de 88 mensagens de pouca ou nenhuma relevância para o processo de identificação de rumores sobre síndrome diarreica em andamento. O resumo geral da classificação das mensagens é exibido na Tabela 9.

Tabela 9 - Classificações - Resumo Geral

Busca por Rumor sobre Síndrome Diarreica		
Tweets	Total	Percentual
Capturados	1941	100,00%
▶ Não Classificados	1808	93,15%
▶ Analisados	133	6,85%
▶ Descartados	88	4,53%
▶ Selecionados	45	2,32%

Resultantes ao processo de descarte, obtivemos 45 mensagens classificadas como “mensagens selecionadas”. Segundo os parâmetros estabelecidos por esta instância do framework, de 1941 mensagens, essas 45 mensagens foram apreciadas pelos especialistas para que estabeleçam a existência ou não de um rumor na região sul do Rio de Janeiro

Segundo os especialistas, as 45 mensagens (2,32%) possuem conteúdo com potencial para direcionar a possível detecção de um rumor, dentro da região determinada para captura de mensagens. A partir desta detecção confirmada, a informação levantada por esta instância fica disponível para a elaboração de possíveis relatórios e emissão de alerta às equipes envolvidas para melhor aprofundamento da investigação.

6.1.3 IDENTIFICAÇÃO DE RUMOR SOBRE SÍNDROME RESPIRATÓRIA

Para a segunda simulação desta pesquisa, foi apresentada uma investigação sobre a possível existência de rumor sobre síndrome respiratória também na Zona Sul do estado do Rio de Janeiro.

Assim como na instância anterior, para estabelecimento das palavras chaves referentes ao grupo de sintomas primários da síndrome respiratória, seus sinônimos e palavras correlacionadas, essa pesquisa totalizou 27 palavras-chaves, conforme exibido na Tabela 10.

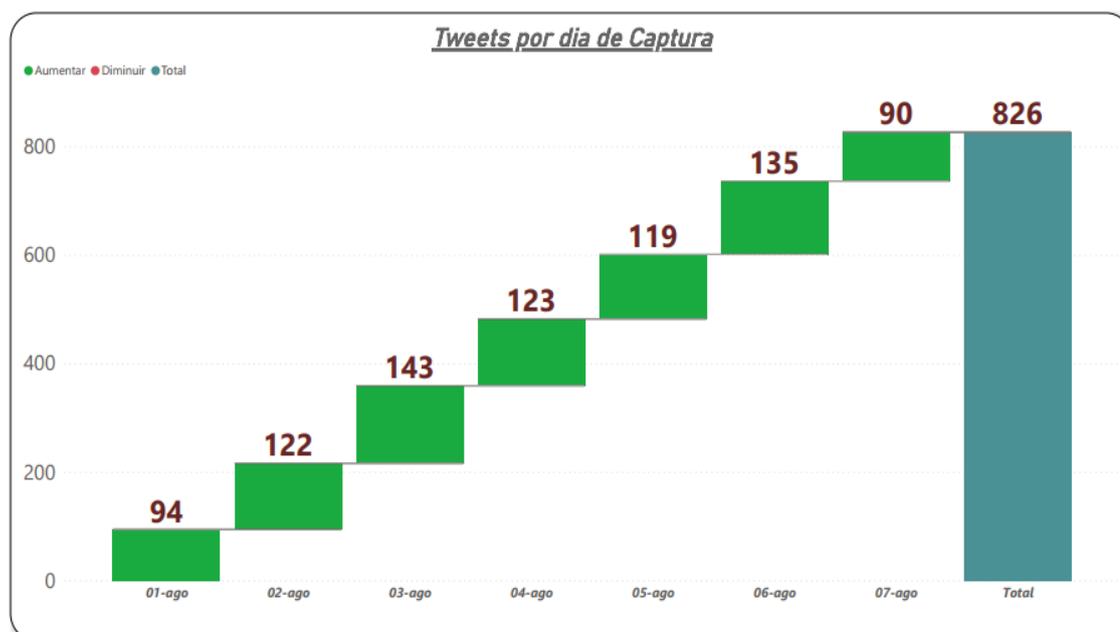
Tabela 10 – Síndrome Respiratória-Plano de Captura

Sinais e Sintomas	tosse	coriza	fadiga
Sinônimos e Variações	expiração	defluxo	estafa
	tossir	muco	nariz entupido
	peitogueira	blenorria	nariz obstruído
	tussícula	dispneia	nariz fechado
	secretar	falta de ar	nariz ocluso
	expelir	abafamento	nariz tapado
	tossinha	sufocação	congestão nasal
	tossicar		asma
	pigarro		

Estabelecidos esses parâmetros para o plano de captura, foi aplicado junto a API do Twitter durante um dia para captura das mensagens correspondentes, compreendidas pelos 7 dias anteriores.

O plano de captura estabelecido para esta instância obteve um total de 826 tweets capturados, distribuídos pelos dias compreendidos no intervalo de tempo conforme mostra o Gráfico 5.

Gráfico 5 – Síndrome Respiratória - Mensagens Capturadas por Dia



Após a captura, as mensagens utilizadas para esta instância foram submetidas a fase de tratamento para identificação daquelas com maior contribuição para identificação de um possível rumor sobre síndrome respiratória. A composição dos parâmetros do plano de análise aplicados a essas 826 mensagens será descrita a seguir:

- ANÁLISE SINDRÔMICA

A primeira etapa é a aplicação da análise sindrômica sobre cada uma das 826 mensagens capturadas. A análise foi realizada através da utilização de uma lógica sindrômica elaborada para esta simulação.

Assim como na instância apresentada anteriormente, esta lógica sindrômica é composta por sintomas considerados secundários para a síndrome respiratória, seus sinônimos e palavras correlacionadas que foram organizadas logicamente através do uso de operadores lógicos. Foram levantadas 27 palavras-chave que foram organizadas

logicamente de maneira mais abrangente com utilização apenas do operador lógico OR, conforme mostrado na Tabela 11.

Tabela 11 - SR - Lógica Sindrômica - Palavras-Chave

Sinais e Sintomas	Febre	OR	Mal Estar	OR	Cefaleia	OR	Dor no Corpo	OR	Dor de Garganta
Sinônimos e Variações	Febril				Cefalia		Algia		garganta arranhada
	Febrento				Cefalgia				Irritação na Garganta
	Febrão				Dor de Cabeça				
	Febrinha				Enxaqueca				
	Febrilizar								
	Febrícula	OR		OR		OR		OR	
	Febriólogo								
	Febroso								
	Hipetermia								
	Hiperpiese								
	hipertermico								
	Pirexia								
	Quentura								
	Temperatura corporal Elevada								
	Temperatura corporal alta								

Após a aplicação da lógica sindrômica apresentada na Tabela 11 sobre as 826 mensagens capturadas, restaram apenas 56 mensagens para continuidade do processo de investigação sobre a existência de rumores sobre a síndrome respiratória.

Nesta instância também é possível observar a eficiência do framework ao reduzir o volume de mensagens que serão analisadas pelas equipes de especialistas. Mesmo que tenham sido coletadas menos mensagens nesta instância do que na anterior, o volume de mensagens ainda é elevado para ser analisada pelos especialistas. A lógica sugerida nesta instância reduziu o volume de 56 mensagens (6,77% do total) que foram identificadas

com maior relação com a síndrome respiratória cuja existência do rumor está sendo investigada.

As 770 mensagens ainda “não classificadas”, ou seja, 93,23% do total, não foram descartadas. Elas representam os tweets que não se enquadraram aos parâmetros referentes à lógica sindrômica aplicada. Novamente, é possível a realização de um refinamento dos parâmetros aplicados a qualquer momento para a possível absorção desses tweets ao novo processo da análise sindrômica.

- **DESCARTE DE MENSAGENS**

Assim como estabelecido no framework e semelhantemente a outra instância desta pesquisa, a seguir serão estabelecidos os parâmetros responsáveis por identificar as mensagens com pouca ou nenhuma contribuição para a investigação realizar a sua classificação como “mensagens descartada”.

Esta instância estabeleceu um total de 20 parâmetros (Tabela 12) para a identificação e classificação das mensagens descartadas para esse processo.

Tabela 12 - Parâmetros de Descarte

Parâmetros de Descarte	
aplicacao	vota
dose	bbb
senado	pared
cachorro	politica
gato	eleicao
bolsonaro	campeonato
presidente	gol
governo	futebol
mandetta	tpm
mutreta	menst

Com aplicação desses parâmetros sobre as 56 mensagens analisadas, foi possível o descarte de 5 mensagens de pouca ou nenhuma relevância para o processo de identificação de rumores sobre síndrome respiratória em andamento. O resumo geral da classificação das mensagens é exibido na Tabela 13.

Tabela 13 - Classificações - Resumo Geral

Busca por Rumor sobre Síndrome Respiratória			
Tweets		Total	Percentual
Capturados		826	100,00%
▶	Não Classificados	770	93,22%
▶	Analisados	56	6,78%
▶	Descartados	5	0,61%
▶	Selecionados	51	6,17%

6.2 ANÁLISE DOS RESULTADOS

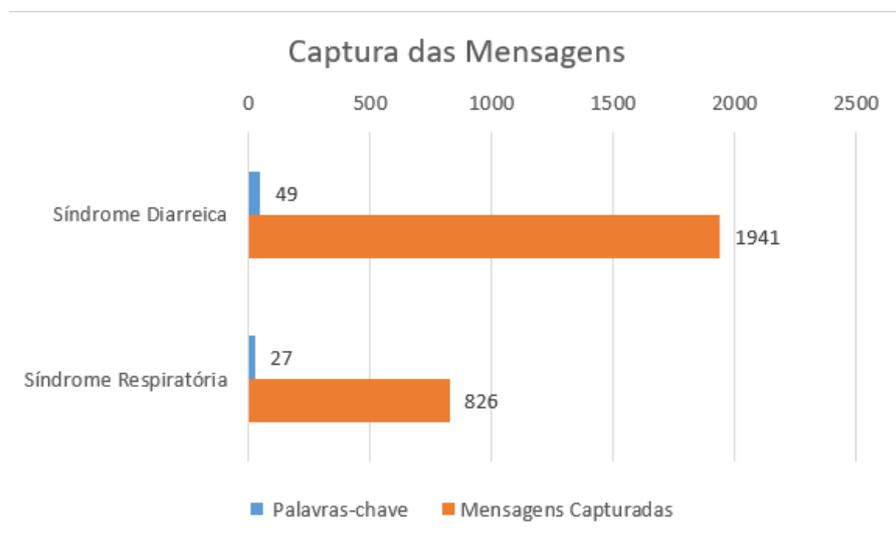
Para efetuar a análise dos resultados obtidos nas instâncias, visando responder à questão de pesquisa deste trabalho, este tópico trata da avaliação da estrutura estabelecida para as simulações e os resultados obtidos.

6.2.1 CAPTURA DAS MENSAGENS

Os parâmetros aplicados foram bastante abrangentes, formados por sintomas primários e seus respectivos sinônimos em cada caso. Essas informações foram obtidas através de uma pesquisa no Google.

A área de abrangência estabelecida e o volume de palavras-chave utilizadas em cada um dos grupos apresentados no Gráfico 6 foi bastante abrangente e contribuiu para obtenção de um maior volume de mensagens, abrangendo maior variabilidade da escrita de um mesmo sintoma primário das respectivas síndromes.

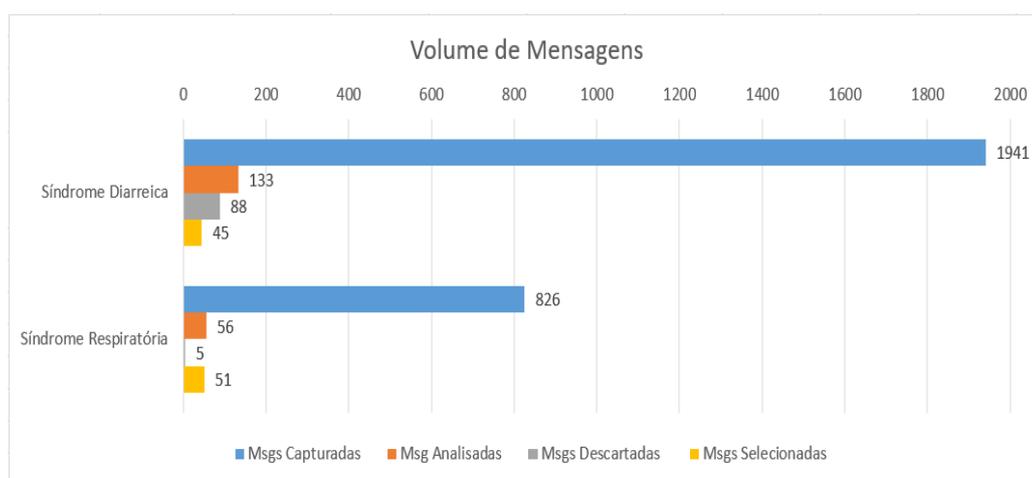
Gráfico 6 - Relação Palavras-Chave e Mensagens Capturadas nas Instâncias



6.2.2 ANÁLISE DAS MENSAGENS

O processo estabelecido pelo Framework e aplicado nas instâncias apresentaram um bom nível de aceitação. Ou seja, a ferramenta auxiliou na automatização da análise sobre as mensagens de Mídias Sociais, identificando entre diversas mensagens, o subgrupo com maior relevância para a investigação segundo os parâmetros aplicados no plano de captura e no plano de análise, conforme mostra o Gráfico 7.

Gráfico 7 - Mensagens das Instâncias



De uma maneira geral, as instâncias elaboradas obtiveram um bom resultado na redução do esforço da análise que seria necessária para abranger o tempo e região geográfica estabelecida nas instâncias.

6.2.3 EMISSÕES DE ALERTAS

As instâncias apresentaram a capacidade do framework em identificar mensagens relevantes ao final de todo o processo que realçaram a necessidade de emissão de alerta sobre a existência de um rumor.

Considerando que as instâncias foram aplicadas em uma área de abrangência correspondente a toda área programática 2.1 no Rio de Janeiro e não apenas em um bairro em específico, os resultados foram satisfatórios.

7 DISCUSSÃO

A motivação para esta pesquisa surgiu através do questionamento sobre como estruturar as informações de Mídias Sociais para ter seu uso facilitado para as equipes de vigilância em saúde pública em seu processo de detecção de rumores sobre síndromes.

A existência de informações valiosas tanto para área da saúde quanto para outras áreas junto às Mídias Sociais foi apurada por diversas pesquisas que também apontaram diversas dificuldades em manipular estas informações para tirar o máximo de utilidade possível. Logo, identificar um processo de agregação, refinamento e geração de relatórios estimulou a elaboração desta pesquisa.

Apesar de diversos estudos e soluções desenvolvidas que buscam utilizar as informações de Mídias Sociais na área da saúde, ainda há algumas oportunidades para novas pesquisas. Em geral, elas não fizeram uso de palavras-chave mais específicas em suas estruturas, como os sintomas de uma mesma síndrome investigada para encontrar informações relevantes (CHEW; EYSENBACH, 2010; DIAZ-AVILES; STEWART, 2012). A pesquisa aqui apresentada se propôs a apresentar uma estrutura de busca através de sintomas dessas possíveis síndromes e seus respectivos sinônimos. Isto possibilitou expandir a qualidade da busca e principalmente a qualidade das mensagens capturadas que irão colaborar para o processo de detecção de rumores.

A pesquisa considera a necessidade da atuação de um especialista durante a realização do processo. Ou seja, é importante a atuação de um especialista na estrutura proposta para identificação de informações relevantes nas mídias sociais para o processo de detecção de rumores em saúde pública para o estabelecimento adequado dos parâmetros a serem aplicados na estrutura da solução.

Mesmo possuindo certo cuidado na identificação das mensagens relevantes nas Mídias Sociais para coletas, algumas pesquisas não apresentaram uma estrutura onde o conhecimento de um especialista pudesse ser explorado, dando-lhes a possibilidade de analisar diferentes resultados através do refinamento de suas análises (ANWAR *et al.*, 2020; DUVAL; SILVA, 2019; WAKAMIYA; KAWAI; ARAMAKI, 2018).

Algumas pesquisas estudadas não apresentaram um tratamento das informações de maneira mais estruturada para identificar informações relevantes (GOMES PESSANHA *et al.*, 2020; SIMÕES; NEVES, 2021), porém este trabalho propôs que a realização do tratamento das informações através da aplicação de uma organização lógica composta de alguns dos sintomas e seus sinônimos relativos a uma mesma síndrome. Durante a

pesquisa foi bastante interessante observar os diferentes resultados obtidos através da aplicação de diferentes lógicas sindrômicas de uma mesma síndrome sobre uma mesma massa de informações já capturadas e como uma lógica pré-existente pode servir de base para construção de outras que irão identificar diferentes síndromes.

Com isso, essa pesquisa elaborou uma estrutura capaz de apoiar as atividades inerentes às equipes de vigilância em saúde. Conforme andamento da pesquisa, as entrevistas realizadas com especialistas e estabelecimento de algumas definições, houve o desenvolvimento de um Framework conceitual para representar os conceitos necessários para o desenvolvimento de soluções tecnológicas que apoiem as equipes de vigilância em saúde em suas atividades de detecção de rumores sobre síndromes.

8 CONCLUSÃO

Esta pesquisa apresentou uma proposta de solução para apoiar o processo de detecção de Rumores sobre Síndromes nas Mídias Sociais. Para isso, foi proposto um framework conceitual para direcionar o desenvolvimento de soluções que consigam lidar com informações capturadas das Mídias Sociais, transformando-as em um recurso útil e oportuno para o processo realizado pelas equipes de vigilância em saúde pública.

O framework descrito, define os parâmetros para captura e análise dessas mensagens, inferindo classificações conforme seu nível de contribuição para a identificação dos rumores. As principais funções foram descritas com suas características e particularidades para o domínio abordado. Houve também a descrição dos principais repositórios de dados necessários e as classificações consideradas para as mensagens. Fluxogramas foram utilizados para enriquecer a descrição de determinadas funções do framework.

Para avaliação do framework, ele foi apresentado em detalhes aos especialistas que participaram desta pesquisa. Em seguida, através de uma entrevista não estruturada, foram realizadas algumas perguntas para identificação de suas críticas e sugestões.

Após a realização dos ajustes apontados na entrevista, e para demonstrar o funcionamento do framework com informações reais, foram elaboradas duas instâncias com informações reais capturadas junto as Mídias Sociais, seguindo os passos descritos no framework proposto nesta pesquisa. Em ambos os momentos o feedback recebido foi satisfatório e agradou aos especialistas que puderam observar seu potencial uso real.

8.1 CONTRIBUIÇÕES

A proposta aborda etapas importantes no suporte de Tecnologia da Informação às equipes de detecção de rumores em saúde pública. Através do framework conceitual proposto nesta pesquisa, foram feitas contribuições importantes para execução deste processo de identificar a existência de rumores sobre síndromes com utilização das informações advindas das Mídias Sociais.

O fluxo de informações, decorrente do processo de captura e análise sugeridos, propõe um mecanismo de classificação das mensagens mais assertivo e direcionado a identificação das informações mais relevantes para o processo.

O armazenamento tanto das mensagens quanto dos parâmetros aplicados em todas as fases do processo abordado pelo framework, representa outra contribuição. Através da organização desses parâmetros sob a forma de planos (captura e análise) é possível maior rastreabilidade e reutilização dos mesmos em diversos momentos. Através dos parâmetros de análise armazenados é possível a aplicação dos refinamentos necessários e a reaplicação dos mesmos sobre as mensagens quantas vezes forem necessárias para atingir o nível de contribuição desejado para conclusão bem-sucedida do processo.

O armazenamento também possibilita a construção de um histórico, com todas as informações utilizadas em cada caso para uso futuro, que permite o registro e aplicação de lições aprendidas e refinamento dos parâmetros em termos de resultados em eventos futuros.

A análise sindrômica proposta estabelece a criação de um repositório criado por especialistas, que organizam logicamente determinados sintomas e sinais da síndrome investigada, formando as lógicas sindrômicas que serão aplicadas a cada uma das mensagens para identificar aquelas de maior relevância e correlação com a investigação.

O fato de estarem armazenadas em um repositório específico, permite sua recuperação sempre que necessário, quer seja para reutilização direta sobre as mensagens, ou como base para criação de outra lógica a ser aplicada. A análise também considera que para uma mesma síndrome pode haver diversas lógicas que podem ser usadas.

8.2 LIMITAÇÕES

A solução proposta por esta pesquisa, ainda não aborda o tratamento de recursos de mídias, ou seja, vídeos, fotos, áudios. Mesmo que possam oferecer consideráveis contribuições às equipes do CIEVS, a limitação tecnológica em atender a complexidade necessária para este tratamento, levou a pesquisa a limitar-se apenas ao tratamento textual.

Em decorrência da falta de tempo hábil, a pesquisa trabalhou com informações advindas apenas do Twitter. Detentora de uma forma de captura mais prática, ela foi abordada em detrimento das demais que para atendimento de suas particularidades requerem maior tempo de estudos e de especificações, para codificação.

Inicialmente não foram consideradas o uso de ontologias ou vocabulários controlados para o framework proposto para classificar as mensagens. Em contato com

especialistas, observou-se que neste momento seria melhor considerar a experiência dos profissionais do CIEVS, onde há carência de soluções tecnológicas para apoiar as atividades de vigilância. A abordagem de estabelecimento de palavras-chaves para estabelecimento de parâmetros do framework, trouxe maior facilidade na compreensão e correlação com a proposta realizada e a contribuição sugerida.

8.3 TRABALHOS FUTUROS

Ferramentas que auxiliem a manipulação de informações das Mídias Sociais, têm sido cada vez mais estudadas e desenvolvidas para oferecer à área da saúde, melhores subsídios para acompanhar determinados cenários epidemiológicos.

Como forma de ampliar a discussão, a apresentação deste framework a especialistas de outras áreas da saúde, especialistas em Mídias Sociais e de Gestão do Conhecimento e tratamento de um grande volume de informações, pode aprimorar ainda mais os conceitos abordados pelo framework. Sua apresentação a especialistas em outras Mídias Sociais diferentes do Twitter também pode trazer uma ampliação significativa da discussão, capaz de contribuir para estudos sobre o funcionamento do Framework sobre outras Mídias Sociais, como Instagram, Facebook etc.

A aplicação de técnicas de processamento de linguagem natural para tratamento das mensagens capturadas pode auxiliar ainda mais na compreensão das mensagens relevantes identificadas pelo framework. A utilização de técnicas mais refinadas para trabalhar com informações de georreferenciamento também pode trazer grande contribuição, ao permitir que as investigações possam ter informações mais detalhadas ao tratarem mensagens oriundas de uma rua específica em não apenas de uma área geográfica específica.

A proposta feita nesta pesquisa apresenta-se sob forma de um framework, mas é possível pensar na integração das soluções elaboradas aos atuais sistemas de vigilância para automatizar ainda mais a complementação das informações desses sistemas com informações advindas das Mídias Sociais.

REFERÊNCIAS

- ABDULLAH, Saeed; WU, Xindong. An Epidemic Model for News Spreading on Twitter. Nov. 2011. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence [...]. [S. l.]: IEEE, Nov. 2011. p. 163–169. <https://doi.org/10.1109/ICTAI.2011.33>.
- ALAMOUDI, A. H.; ZAIDAN, B. B.; ZAIDAN, A. A.; ALBAHRI, O. S.; MOHAMMED, K. I.; MALIK, R. Q.; ALMAHDI, E. M.; CHYAD, M. A.; TAREQ, Z.; ALBAHRI, A. S.; HAMEED, Hamsa; ALAA, Musaab. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, vol. 167, p. 114155, 2021. <https://doi.org/10.1016/j.eswa.2020.114155>.
- ANWAR, Mohd; KHOURY, Dalia; ALDRIDGE, Arnie P.; PARKER, Stephanie J.; CONWAY, Kevin P. Using Twitter to Surveil the Opioid Epidemic in North Carolina: An Exploratory Study. *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. 1–7, 2020. <https://doi.org/10.2196/17574>.
- BAKER, Qanita Bani; SHATNAWI, Farah; RAWASHDEH, Saif; AL-SMADI, Mohammad; JARARWEH, Yaser. Detecting epidemic diseases using sentiment analysis of arabic tweets. *Journal of Universal Computer Science*, vol. 26, no. 1, p. 50–70, 2020. <https://doi.org/10.3897/jucs.2020.004>.
- BARBOZA, Philippe; VAILLANT, Laetitia; MAWUDEKU, Abla; NELSON, Noele P.; HARTLEY, David M.; MADOFF, Lawrence C.; LINGE, Jens P.; COLLIER, Nigel; BROWNSTEIN, John S.; YANGARBER, Roman; ASTAGNEAU, Pascal. Evaluation of Epidemic Intelligence Systems Integrated in the Early Alerting and Reporting Project for the Detection of A/H5N1 Influenza Events. *PLoS ONE*, vol. 8, no. 3, p. e57252, 5 Mar. 2013. <https://doi.org/10.1371/journal.pone.0057252>.
- BEZERRA, Anselmo César Vasconcelos; BITOUN, Jan. Metodologia participativa como instrumento para a territorialização das ações da vigilância em saúde ambiental. *Ciência e Saúde Coletiva*, vol. 22, no. 10, p. 3259–3268, 2017. <https://doi.org/10.1590/1413-812320172210.17722017>.
- BOWSER, G.; MILNER, C.; SULLIVAN, R. Medical intelligence, security and global health: the foundations of a new health agenda. *Journal of the Royal Society of Medicine*, vol. 109, no. 7, p. 269–273, 6 Jul. 2016. <https://doi.org/10.1177/0141076816656483>.
- BOYD, danah m.; ELLISON, Nicole B. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, vol. 13, no. 1, p. 210–230, Oct. 2007. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>.
- BRASIL; AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA; ORGANIZAÇÃO MUNDIAL DA SAÚDE. Regulamento Sanitário Internacional. [S. l.: s. n.], 2005.
- BRASIL, Ministério da Saúde. Novo coronavírus (2019-nCoV). *Boletim Epidemiológico*, vol. 51, no. 4, 2020.
- BRASIL. MINISTÉRIO DA SAÚDE. Os CIEVS são referências nacionais na detecção e na resposta às emergências. 2015. Ministério da Saúde.

BRASIL.; MINISTÉRIO DA SAÚDE; SAÚDE; EM, Secretaria de Vigilância. Centro de informações estratégicas em vigilância em saúde - CIEVS. Brasília / DF, p. 11, 2007.

BRASIL; MINISTÉRIO DA SAÚDE; SECRETARIA MUNICIPAL DE SAÚDE. COVID-19. Boletim Epidemiológico, vol. 28, no. 1, p. 25, Apr. 2021.

BRASIL.; SAÚDE., Ministério da; SAÚDE., Secretaria de Vigilância em; TRANSMISSÍVEIS, Departamento de Vigilância das Doenças. Plano de operação do ponto focal nacional para o regulamento sanitário internacional. Brasília, DF: Ministério da Saúde, 2016.

BRASIL; SECRETARIA DE VIGILÂNCIA EM SAÚDE; CIEVS. Plano de Resposta a Emergências em Saúde Pública. 2018. CIEVS-Divisão de Informação. Available at: <http://www.saude.pr.gov.br/modules/conteudo/conteudo.php?conteudo=3075>. Accessed on: 1 Nov. 2019. (Citado na Dissertação como CIEVS 2018).

CARBINATTO, Bruno. China diz que wet market de Wuhan provavelmente não foi origem da Covid-19. 2020. Super Interessante. Available at: <https://super.abril.com.br/saude/china-diz-que-wet-market-de-wuhan-provavelmente-nao-foi-origem-da-covid-19/>. Accessed on: 2 Jun. 2020.

CAREGNATO, Rita Catalina Aquino. Pesquisa qualitativa. Rev. SOBECC, p. 1–2, 2017.

CARMO, Eduardo Hage. Emergências de saúde pública: breve histórico, conceitos e aplicações. Saúde em Debate, vol. 44, no. spe2, p. 9–19, Jul. 2020. <https://doi.org/10.1590/0103-11042020e201>.

CHANDRAPAUL; SONI, Rounak; SHARMA, Shubham; FAGNA, Hemraj; MITTAL, Sangeeta. News analysis using word cloud. [S. l.]: Springer Singapore, 2019. vol. 526,. https://doi.org/10.1007/978-981-13-2553-3_6.

CHEW, Cynthia; EYSENBACH, Gunther. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. PLoS ONE, vol. 5, no. 11, p. e14118, 29 Nov. 2010. <https://doi.org/10.1371/journal.pone.0014118>.

CHUNARA, Rumi; ANDREWS, Jason R.; BROWNSTEIN, John S. Social and News Media Enable Estimation of Epidemiological CHUNARA, R.; ANDREWS, J. R.; BROWNSTEIN, J. S. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. The American Journal of Tropical Medic. The American Journal of Tropical Medicine and Hygiene, vol. 86, no. 1, p. 39–45, 1 Jan. 2012. <https://doi.org/10.4269/ajtmh.2012.11-0597>.

CRESWELL, J W. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. 3rd ed. CA: SAGE Publications, 2012.

DAKKAK, H.; BROWN, R.; TWYNSTRA, J.; CHARBONNEAU, K.; SEABROOK, J. A. The perception of pre-and post-natal marijuana exposure on health outcomes: A content analysis of Twitter messages. Journal of Neonatal-Perinatal Medicine, vol. 11, no. 4, p. 409–415, 2018. <https://doi.org/10.3233/NPM-17133>.

DENECKE, Kerstin. An ethical assessment model for digital disease detection technologies. *Life Sciences, Society and Policy*, vol. 13, no. 1, 2017. <https://doi.org/10.1186/s40504-017-0062-x>.

DIAS, Donaldo De Souza; SILVA, Mônica Ferreira da. *COMO ESCREVER UMA MONOGRAFIA*. [S. l.: s. n.], 2009.

DIAZ-AVILES, Ernesto; STEWART, Avaré. Tracking Twitter for Epidemic Intelligence. Case study: EHEC/HUS outbreak in Germany, 2011. *Proceedings of the 4th Annual ACM Web Science Conference, WebSci'12*, vol. volume, p. 82–85, 2012. <https://doi.org/10.1145/2380718.2380730>.

DIAZ-AVILES, Ernesto; STEWART, Avaré; VELASCO, Edward; DENECKE, Kerstin; NEJDL, Wolfgang. Epidemic intelligence for the crowd, by the crowd. *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, no. *Icws*, p. 439–442, 2012.

DIFONZO, Nicholas; BORDIA, Prashant. Rumor, gossip and urban legends. *Diogenes*, vol. 54, no. 1, p. 19–35, 2007. <https://doi.org/10.1177/0392192107073433>.

DODGE, Graham. Using social media as a public health surveillance. 2017. *Becker's Hospital Review*. Available at: <https://www.beckershospitalreview.com/care-coordination/using-social-media-as-a-public-health-surveillance-tool.html>. Accessed on: 1 Apr. 2020.

DUVAL, Felipe Vieira; SILVA, Fabrício Alves Barbosa da. O uso do Twitter como minerador de eventos adversos de medicamentos de combate à malária: o caso da doxiciclina. *Cadernos de Saúde Pública*, vol. 35, no. 5, p. 1–16, 2019. <https://doi.org/10.1590/0102-311x00033417>.

EYSENBACH, Gunther. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA ... Annual Symposium proceedings / AMIA Symposium*. *AMIA Symposium*, p. 244–248, 2006.

FERGUSON, Neil M.; CUMMINGS, Derek A.T.; CAUCHEMEZ, Simon; FRASER, Christophe; RILEY, Steven; MEEYAI, Aronrag; IAMSIRITHAWORN, Soapon; BURKE, Donald S. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, vol. 437, no. 7056, p. 209–214, 3 Sep. 2005. <https://doi.org/10.1038/nature04017>.

FERREIRA, Fátima de Paula; CASTRO, Larissa de Paula Gonzaga e. O NOVO REGULAMENTO SANITÁRIO INTERNACIONAL E O CONTROLE DA PANDEMIA INFLUENZA H1N1 NO BRASIL. *Revista Paradigma*, vol. 20, no. 5, p. 255, 2011.

FORTES, Paulo Antônio de Carvalho; RIBEIRO, Helena. Saúde Global em tempos de globalização. *Saúde e Sociedade*, vol. 23, no. 2, p. 366–375, Jun. 2014. <https://doi.org/10.1590/S0104-12902014000200002>.

FREITAS, Danilo P.; BORGES, Marcos R. S.; CARVALHO, Paulo Victor R. de. A conceptual framework for developing solutions that organise social media information for

emergency response teams. *Behaviour & Information Technology*, vol. 39, no. 3, p. 360–378, 3 Mar. 2019. <https://doi.org/10.1080/0144929X.2019.1621933>.

FREITAS, Danielle Amaral De; SOUZA-SANTOS, Reinaldo; WAKIMOTO, Mayumi Duarte. Acesso aos serviços de saúde por pacientes com suspeita de dengue na cidade do Rio de Janeiro, Brasil. *Ciência & Saúde Coletiva*, vol. 24, no. 4, p. 1507–1516, Apr. 2019. <https://doi.org/10.1590/1413-81232018244.11252017>.

GIL, Antonio Carlos. *Métodos e Técnicas de Pesquisa Social*. [S. l.: s. n.], 2008. vol. 264,.

GIUSTINI, Dean. How Web 2.0 is changing medicine. *BMJ*, vol. 333, no. 7582, p. 1283–1284, 23 Dec. 2006. <https://doi.org/10.1136/bmj.39062.555405.80>.

GOMES PESSANHA, Gabriel Rodrigo; OLIVEIRA FIDELIS, Taylor; DOURADO FREIRE, Clarissa; ALMEIDA SOARES, Eduardo. #Fiqueemcasa: Análise De Sentimento Dos Usuários Do Twitter Em Relação Ao Covid19. *Holos*, vol. 5, p. 1–20, 2020. <https://doi.org/10.15628/holos.2020.11147>.

GREENLEAF, Abigail R.; MWIMA, Gerald; LETHOKO, Molibeli; CONKLING, Martha; KEEFER, George; CHANG, Christiana; MCLEOD, Natasha; MARUYAMA, Haruka; CHEN, Qixuan; FARLEY, Shannon M.; LOW, Andrea. Participatory surveillance of COVID-19 in lesotho via weekly calls: Protocol for cell phone data collection. *JMIR Research Protocols*, vol. 10, no. 9, 2021. <https://doi.org/10.2196/31236>.

GREGOR, Shirley; HEVNER, Alan R. Positioning and presenting design science research for maximum impact. *MIS Quarterly: Management Information Systems*, vol. 37, no. 2, p. 337–355, 2013. <https://doi.org/10.25300/MISQ/2013/37.2.01>.

GRISOTTI, Márcia. Doenças infecciosas emergentes e a emergência das doenças: uma revisão conceitual e novas questões. *Ciência & Saúde Coletiva*, vol. 15, no. suppl 1, p. 1095–1104, 2010. <https://doi.org/10.1590/s1413-81232010000700017>.

GUI, Xinning; KOU, Yubo; PINE, Kathleen H.; CHEN, Yunan. Managing uncertainty: Using social media for risk assessment during a public health crisis. *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, p. 4520–4533, 2017. <https://doi.org/10.1145/3025453.3025891>.

HAGEN, Loni; NEELY, Stephen; SCHARF, Ryan; KELLER, Thomas E. Social Media Use for Crisis and Emergency Risk Communications During the Zika Health Crisis. *Digital Government: Research and Practice*, vol. 1, no. 2, p. 1–21, 17 Apr. 2020. <https://doi.org/10.1145/3372021>.

HU, Xia; TANG, Lei; LIU, Huan. Enhancing accessibility of microblogging messages using semantic knowledge. 2011. *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11 [...]*. New York, New York, USA: ACM Press, 2011. p. 2465. <https://doi.org/10.1145/2063576.2063993>.

JAIN, Vinay Kumar; KUMAR, Shishir. An Effective Approach to Track Levels of Influenza-A (H1N1) Pandemic in India Using Twitter. *Procedia Computer Science*, vol. 70, p. 801–807, 2015. <https://doi.org/10.1016/j.procs.2015.10.120>.

JONES, Nickolas M.; THOMPSON, Rebecca R.; SCHETTER, Christine Dunkel; SILVER, Roxane Cohen. Distress and rumor exposure on social media during a campus lockdown. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 44, p. 11663–11668, 2017. <https://doi.org/10.1073/pnas.1708518114>.

KAPLAN, Andreas M.; HAENLEIN, Michael. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, vol. 53, no. 1, p. 59–68, Jan. 2010. <https://doi.org/10.1016/j.bushor.2009.09.003>.

KASS-HOUT, Taha A.; ALHINNAWI, Hend. Social media in public health. *British Medical Bulletin*, vol. 108, no. 1, p. 5–24, 1 Dec. 2013. <https://doi.org/10.1093/bmb/ldt028>.

KLEIN, Gisiela Hasse; GUIDI NETO, Pedro; TEZZA, Rafael. Big Data e mídias sociais: monitoramento das redes como ferramenta de gestão. *Saúde e Sociedade*, vol. 26, no. 1, p. 208–217, Mar. 2017. <https://doi.org/10.1590/s0104-12902017164943>.

LAI, Xiaoquan; WANG, Minghuan; QIN, Chuan; TAN, Li; RAN, Lusen; CHEN, Daiqi; ZHANG, Han; SHANG, Ke; XIA, Chen; WANG, Shaokang; XU, Shabei; WANG, Wei. Coronavirus Disease 2019 (COVID-2019) Infection Among Health Care Workers and Implications for Prevention Measures in a Tertiary Hospital in Wuhan, China. *JAMA Network Open*, vol. 3, no. 5, p. e209666, 21 May 2020. <https://doi.org/10.1001/jamanetworkopen.2020.9666>.

LEAL-NETO, Onício B.; DIMECH, George S.; LIBEL, Marlo; OLIVEIRA, Wanderson; FERREIRA, Juliana Perazzo. Detecção digital de doenças e vigilância participativa: panorama e perspectivas para o Brasil. *Revista de Saúde Pública*, vol. 50, p. 17, 2016. <https://doi.org/10.1590/S1518-8787.2016050006201>.

LIANG, Hai; FUNG, Isaac Chun Hai; TSE, Zion Tsz Ho; YIN, Jingjing; CHAN, Chung Hong; PECHTA, Laura E.; SMITH, Belinda J.; MARQUEZ-LAMEDA, Rossmary D.; MELTZER, Martin I.; LUBELL, Keri M.; FU, King Wa. How did Ebola information spread on twitter: Broadcasting or viral spreading? *BMC Public Health*, vol. 19, no. 1, p. 1–11, 2019. <https://doi.org/10.1186/s12889-019-6747-8>.

LIAO, Qiuyan; YUAN, Jiehu; DONG, Meihong; YANG, Lin; FIELDING, Richard; LAM, Wendy Wing Tak. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in china: Infodemiology study on social media data. *Journal of Medical Internet Research*, vol. 22, no. 5, p. 1–13, 2020. <https://doi.org/10.2196/18796>.

LIMA, Trindade. O Brasil e a Organização Pan-Americana da Saúde: uma história em três dimensões. In: JACOBO FINKELMAN (ed.). *Caminhos da Saúde Pública no Brasil*. Rio de Janeiro: Fiocruz, 2002. p. 23–116.

LONGINI, Ira M. Containing Pandemic Influenza at the Source. *Science*, vol. 309, no. 5737, p. 1083–1087, 12 Aug. 2005. <https://doi.org/10.1126/science.1115717>.

LWIN, May Oo; LU, Jiahui; SHELDENKAR, Anita; SCHULZ, Peter Johannes; SHIN, Wonsun; GUPTA, Raj; YANG, Yinping. Global sentiments surrounding the COVID-19

pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. 1–4, 2020. <https://doi.org/10.2196/19447>.

MARKIEWICZ, Marcus E; LUCENA, Carlos. Understanding Object-Oriented Framework Engineering. *PUC-RioInf.MCC38*, p. 11, 2000.

MATOS, Vanina. A saúde do viajante na visão de três atores: gestores da saúde pública, gestores do turismo e o turista. 2011. 132 f. Escola Nacional de Saúde Pública Sergio Arouca, 2011.

MATTSSON, Michael. Evolution and Composition of Object-Oriented Frameworks. 2000. 216 f. Blekinge Institute of Technology, 2000.

MAVRAGANI, Amaryllis. Infodemiology and infoveillance: Scoping review. *Journal of Medical Internet Research*, vol. 22, no. 4, p. e16206, 28 Apr. 2020. <https://doi.org/10.2196/16206>.

MILES, Matthew B Mathew B.; HUBERMAN, A.M. Qualitative data analysis: an expanded sourcebook / Matthew B. Miles, A. Michael Huberman. 2nd ed. Thousand Oaks: Sage Publications, 1994(, book).

MINAYO, Maria Cecília de Souza. O desafio do conhecimento: pesquisa qualitativa em saúde / Maria Cecília de Souza Minayo., p. 407, 2010.

MINISTÉRIO DA SAÚDE; SECRETARIA DE VIGILÂNCIA EM SAÚDE. Boletins Epidemiológicos. 2022. Ministério da Saúde. Available at: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos>. Accessed on: 31 Jan. 2022.

MINISTÉRIO DA SAÚDE; SECRETARIA DE VIGILÂNCIA EM SAÚDE. Doenças Tropicais Negligenciadas. *Boletim Epidemiológico*, vol. Especial, no. 1, p. 76, 2021.

MONDAL, Tamal; ROY, Tunir; BHATTACHARYA, Indrajit; BHATTACHARYA, Sourav; DAS, Indranil. A study on rumor propagation trends and features in a post disaster situation. *ACM International Conference Proceeding Series*, p. 381–384, 2019. <https://doi.org/10.1145/3288599.3295581>.

MOONEY, Stephen J.; PEJAVER, Vikas. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annual Review of Public Health*, vol. 39, p. 95–112, 2018. <https://doi.org/10.1146/annurev-publhealth-040617-014208>.

MOORHEAD, S. Anne; HAZLETT, Diane E.; HARRISON, Laura; CARROLL, Jennifer K.; IRWIN, Anthea; HOVING, Ciska. A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, vol. 15, no. 4, p. e85, 23 Apr. 2013. <https://doi.org/10.2196/jmir.1933>.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. Folha informativa – COVID-19 (doença causada pelo novo coronavírus). 2020. Folha Informativa - COVID19. Available at:

https://www.paho.org/bra/index.php?option=com_content&view=article&id=6101:covid19&Itemid=875. Accessed on: 5 Apr. 2020.

OZTURK, Pinar; LI, Huaye; SAKAMOTO, Yasuaki. Combating rumor spread on social media: The effectiveness of refutation and warning. *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2015-March, p. 2406–2414, 2015. <https://doi.org/10.1109/HICSS.2015.288>.

PAQUET, C.; COULOMBIER, D.; KAISER, R.; CIOTTI, M. Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. *Eurosurveillance*, vol. 11, no. 12, p. 5–6, 1 Dec. 2006. <https://doi.org/10.2807/esm.11.12.00665-en>.

PAUL, Michael J; DREDZE, Mark. You Are What You Tweet: Analyzing Twitter for Public Health. *Int’L AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.

PAUL, Michael J.; SARKER, Abeed; BROWNSTEIN, John S.; NIKFARJAM, Azadeh; SCOTCH, Matthew; SMITH, Karen L.; GONZALEZ, Graciela. Social media mining for public health monitoring and surveillance. *Pacific Symposium on Biocomputing*, p. 468–479, 2016. https://doi.org/10.1142/9789814749411_0043.

PEREZ, Beatrice; MUSOLESI, Mirco; STRINGHINI, Gianluca. You Are Your Metadata: Identification and Obfuscation of Social Media Users Using Metadata Information. *no. Icwsm*, p. 241–250, 2018.

PIMENTEL, Renata Macedo Martins; DABOIN, Blanca Elena Guerrero; OLIVEIRA, Adriana Gonçalves de; JR, Hugo Macedo Jr. A disseminação da covid-19: um papel expectante e preventivo na saúde global. *J Hum Growth*, vol. 30, no. 1, p. 135–140, 2020.

PLATT, Alana; CITRIN, Levi; HOOD, Cynthia. Healthshark: Using Twitter for Situational Awareness in Public Health. *Journal of Information Privacy and Security*, vol. 11, no. 1, p. 19–37, 2 Jan. 2015. <https://doi.org/10.1080/15536548.2015.1010984>.

PRIETO-DÍAZ, Rubén. Domain analysis. *ACM SIGSOFT Software Engineering Notes*, vol. 15, no. 2, p. 47–54, Apr. 1990. <https://doi.org/10.1145/382296.382703>.

ROVETTA, Alessandro; BHAGAVATHULA, Akshaya Srikanth. Global infodemiology of COVID-19: Analysis of Google web searches and Instagram hashtags. *Journal of Medical Internet Research*, vol. 22, no. 8, p. 1–9, 2020. <https://doi.org/10.2196/20673>.

SANTOS, José Carlos; MATOS, Sérgio. Analysing Twitter and web queries for flu trend prediction. *Theoretical Biology and Medical Modelling*, vol. 11, no. SUPPL.1, p. S6, 2014. <https://doi.org/10.1186/1742-4682-11-S1-S6>.

SECRETARIA DE VIGILÂNCIA EM SAÚDE; MINISTÉRIO DA SAÚDE. Plano de Resposta às Emergências em saúde pública. 1st ed. Brasília/DF: [s. n.], 2014a. vol. 2018-02–28.

SECRETARIA DE VIGILÂNCIA EM SAÚDE; MINISTÉRIO DA SAÚDE. Plano de Resposta às Emergências em saúde pública. 1st ed. Brasília/DF: [s. n.], 2014b. Vol. 2018-02–28.

SEIN, Maung K.; HENFRIDSSON, Ola; PURAO, Sandeep; ROSSI, Matti; LINDGREN, Rikard. Action design research. *MIS Quarterly: Management Information Systems*, vol. 35, no. 1, p. 37–56, 2011. <https://doi.org/10.2307/23043488>.

SIGNORINI, Alessio. Social Web Information Monitoring for Health. Department of Computer Science University of Iowa, 7 jul. 2009.

SILVA, Francisca Jocineide da Costa; CARVALHO, Maria Eulina Pessoa de. O Estado Da Arte Das Pesquisas Educacionais Sobre Gênero E Educação Infantil: Uma Introdução. 18 Redor, p. 346–362, 2014.

SIMÕES, Maria Denise; NEVES, Ana Régia de M. Mineração de Texto no Twitter: uma ferramenta auxiliar na detecção de epidemias, p. 281–292, 2021. <https://doi.org/10.5753/sbcas.2021.16072>.

SIMON, H A; LAIRD, J E. *The Sciences of the Artificial*. [S. l.]: MIT Press, 2019 (The MIT Press).

SOUZA, Cleidson Ronald Botelho de. Um Framework para Editores de Diagramas Cooperativos baseados em Anotações. 1998.

STIEGLITZ, Stefan; MIRBABAIE, Milad; ROSS, Björn; NEUBERGER, Christoph. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, vol. 39, no. October 2017, p. 156–168, 2018. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>.

TEIXEIRA, Maria Glória; COSTA, Maria da Conceição Nascimento; CARMO, Eduardo Hage; OLIVEIRA, Wanderson Kleber de; PENNA, Gerson Oliveira. Vigilância em Saúde no SUS - construção, efeitos e perspectivas. *Ciência & Saúde Coletiva*, vol. 23, no. 6, p. 1811–1818, Jun. 2018. <https://doi.org/10.1590/1413-81232018236.09032018>.

TEIXEIRA, Maria Glória; COSTA, Maria da Conceição Nascimento; PAIXÃO, Enny Santos da; CARMO, Eduardo Hage; BARRETO, Florisneide Rodrigues; PENNA, Gerson Oliveira. Conquistas do SUS no enfrentamento das doenças transmissíveis. *Ciência & Saúde Coletiva*, vol. 23, no. 6, p. 1819–1828, Jun. 2018. <https://doi.org/10.1590/1413-81232018236.08402018>.

THE GUARDIAN. Cities after coronavirus: how Covid-19 could radically alter urban life. 2020. Available at: https://www.theguardian.com/world/2020/mar/26/life-after-coronavirus-pandemic-change-world?CMP=Share_iOSApp_Other.

TWEET. TweetDeck. 2022. Available at: <https://tweetdeck.twitter.com/>. Accessed on: 9 Jan. 2022.

UNIVERSITY, Johns Hopkins. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE). 2021. Available at: <https://ourworldindata.org/explorers/coronavirus-data-explorer?tab=map&yScale=log&zoomToSelection=true&hideControls=true&Metric=Confirmed+cases&Interval=Cumulative&Relative+to+Population=false&Align+outbreaks=true>. Accessed on: 2 Jun. 2021.

VERGARA, Sylvia Constant. *Começando a definir a metodologia*. 4th ed. São Paulo: [s. n.], 2003.

VERMA, Sudha; VIEWEG, Sarah; CORVEY, William J; PALEN, Leysia; MARTIN, James H; PALMER, Martha; SCHRAM, Aaron; ANDERSON, Kenneth M. Natural Language Processing to the Rescue? Extracting ``Situational Awareness`` Tweets During Mass Emergency. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, no. June 2014, p. 385–392, 2011.

VEYSEH, Amir Poursan Ben; EBRAHIMI, Javid; DOU, Dejing; LOWD, Daniel. A Temporal Attentional Model for Rumor Stance Classification. Part F1318., 6 Nov. 2017. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management [...]*. New York, NY, USA: ACM, 6 Nov. 2017. vol. Part F1318, p. 2335–2338. <https://doi.org/10.1145/3132847.3133116>.

VILAIN, Pascal; MENUJIER, Luce; FILLEUL, Laurent. Twitter: a complementary tool to monitor seasonal influenza epidemic in France? *Online Journal of Public Health Informatics*, vol. 11, no. 1, p. 2017–2020, 2019. <https://doi.org/10.5210/ojphi.v11i1.9724>.

WAKAMIYA, Shoko; KAWAI, Yukiko; ARAMAKI, Eiji. Twitter-based influenza detection after flu peak via tweets with indirect information: Text mining study. *JMIR Public Health and Surveillance*, vol. 4, no. 9, p. 1–17, 2018. <https://doi.org/10.2196/publichealth.8627>.

WANG, Huifeng; GAN, Jianhou; CHEN, Jianbin; OUYANG, Zhaoxiang. Automatic detecting for COVID-19-related rumors data on internet. *ACM International Conference Proceeding Series*, p. 22–26, 2021. <https://doi.org/10.1145/3456415.3456420>.

WHO. *International Health Regulations*, n. 3. Geneva: WHO Press, 2005.

WORLD HEALTH ORGANIZATION. *Early detection, assessment and response to acute public health events*. [S. l.: s. n.], 2014. vol. Interim,.

WORLD HEALTH ORGANIZATION. *WHA58.3 Revision of the International Health Regulations: Resolutions and Decisions*. 58., 2005. *World Health Assembly 58 (2005) Official Records [...]*. Geneva: [s. n.], 2005. vol. 58, p. 57.

XAVIER, Fernando; OLENSCKI, João Rodrigo W.; ACOSTA, Andre Luis; SALLUM, Maria Anice Mureb; SARAIVA, Antonio Mauro. Análise de redes sociais como estratégia de apoio a vigilância em saúde durante a Covid-19. *Estudos Avançados*, vol. 34, no. 99, p. 261–282, 2020. <https://doi.org/10.1590/S0103-4014.2020.3499.016>.

YANG, Y. Tony; HORNEFFER, Michael; DILISIO, Nicole. Mining social media and web searches for disease detection. *Journal of Public Health Research*, vol. 2, no. 1, p. 4, 31 May 2013. <https://doi.org/10.4081/jphr.2013.e4>.

YAN, S. J.; CHUGHTAI, A. A.; MACINTYRE, C. R. Utility and potential of rapid epidemic intelligence from internet-based sources. *International Journal of Infectious Diseases*, vol. 63, p. 77–87, 2017. <https://doi.org/10.1016/j.ijid.2017.07.020>.

ZHANG, Zili; ZHANG, Ziqiong; LI, Hengyun. Predictors of the authenticity of Internet health rumours. *Health Information & Libraries Journal*, vol. 32, no. 3, p. 195–205, Sep. 2015. <https://doi.org/10.1111/hir.12115>.

ZHOU, Xichuan; YE, Jieping; FENG, Yujie. Tuberculosis surveillance by analyzing google trends. *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 8, p. 2247–2254, 2011. <https://doi.org/10.1109/TBME.2011.2132132>.



Instituto Tércio Pacitti de
Aplicações e Pesquisas
Computacionais



Instituto de Matemática



1920 | 2020

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Av. Athos da Silveira Ramos, 274
Prédio do CCMN – Cidade Universitária
Ilha do Fundão – Rio de Janeiro
21941-916
Brasil

