

A disseminação de informações associadas a epidemia do Zika Vírus: uma pesquisa a partir de dados coletados no Twitter

Guilherme Ataíde Dias^a, Denyssson Axel Ribeiro Mota^b, André Luiz Dias de França^c e Eduardo de Santana Medeiros Alexandre^d

Resumo: A pesquisa apresentada, no momento em andamento, investiga de que maneira a Ciência da Informação pode fornecer suporte para entendimento de questões que emergem a partir da disseminação de informações acerca do Zika vírus postadas na rede social Twitter. O objetivo principal é investigar a disseminação de informações através de canais formais e informais como redutora de incertezas de uma população imersa em epidemia de Zika Vírus. A pesquisa configura-se com relação aos procedimentos técnicos como bibliográfica, com relação aos objetivos é exploratória e possui natureza quantitativa e qualitativa. O corpus estudado constituiu-se de mensagens disseminadas (tweet) através da rede social Twitter com conteúdos relacionados ao Zika Vírus no período compreendido entre outubro de 2017 até março de 2018. A partir dos dados coletados é possível mencionar que os usuários do Twitter que postam mensagens com o termo “zica” e strings equivalentes, não estão na maioria das vezes fazendo referência às questões de saúde associadas com a epidemia de Zika Vírus. O termo zica (zika) é usado em uma grande variedade de contextos que transcendem o domínio específico da área da saúde. A continuidade da pesquisa possibilitará uma compreensão mais aprofundada do fenômeno.

Palavras-chave: Twitter, Disseminação de dados, Zika vírus.

The dissemination of information associated with the zika virus epidemic: A research based on data collected from the Twitter

Abstract: The ongoing research presented investigates how Information Science can provide support for understanding issues that emerge from the dissemination of information about the Zika virus posted on the social network Twitter. The main objective is to investigate the dissemination of information through formal and informal channels as an element capable of reducing of uncertainties on a population immersed in a Zika Virus epidemic. The research is characterized in relation to the technical procedures as being bibliographic, with regard to the objectives is exploratory and has a quantitative and qualitative nature. The studied corpus consisted of tweet messages disseminated through the social network Twitter with contents related to the Zika Virus in the period between October 2017 until March 2018. From the data collected it is possible to mention that Twitter users who post messages with the term "zica" and equivalent strings are not often referring to the health issues associated with the Zika Virus epidemic. The term zica (zika) is used in a wide variety of contexts that transcend the specific domain of health. The continuity of the research will allow a deeper understanding of the phenomenon.

Keyword: Twitter; Data dissemination; Zika virus.

-
- a Universidade Federal da Paraíba (UFPB). E-mail: guilhermeataide@ccsa.ufpb.br. ORCID: <https://orcid.org/0000-0001-6576-0017>. Currículo: <http://lattes.cnpq.br/9553707435669429>
 - b Universidade Federal da Paraíba (UFPB). E-mail: denyssson.mota@ufca.edu.br. ORCID: <https://orcid.org/0000-0003-2069-134X>. Currículo: <http://lattes.cnpq.br/1273037072033521>
 - c Universidade Federal da Paraíba (UFPB). E-mail: andreluizjp@gmail.com. ORCID: <https://orcid.org/0000-0002-8278-3664>. Currículo: <http://lattes.cnpq.br/5684942906411942>
 - d Instituto Federal da Paraíba (IFPB). E-mail: eduardo.ufpb@gmail.com. Currículo: <http://lattes.cnpq.br/6047383716536661>

1 Introdução⁷³

Vários temas relacionados com a área da saúde possuem ligação com a Ciência da Informação. Exemplos de trabalhos que referendam este pensamento podem ser encontrados nos materiais organizados por Duarte e Farias (2009), Galvão e Ricarte (2012) e Pinto e Campos (2013).

De forma mais ampla ilustramos a essência das intersecções entre a Ciência da Informação e a área da saúde através do discurso de Gomez (DIAS, 2016):

A convergência de questões de informação e questões em saúde é de longa data, sendo a área de informação em saúde rica e pioneira na inovação e pesquisa em informação e em tecnologias de informação. A otimização e atualização de dispositivos e ações de informação têm de responder às demandas intensivas de informação e comunicação que satisfação, em quantidade e qualidade, às metas e finalidades da promoção da saúde e do atendimento clínico. A Ciência da Informação compõe uma das muitas redes inter e transdisciplinares onde essas questões e temáticas são objeto de pesquisa e reflexão (GOMEZ, 2013, p.29).

É de amplo conhecimento que o mundo, e de forma mais marcante os países da América Central e do Sul estão sendo atingidos por uma série de enfermidades atribuídas ao Zika Vírus (Vide Figura 1), como por exemplo, a microcefalia e a Síndrome de Guillain– Barré (DIAS, 2016). Este é um exemplo de fenômeno que pode ser estudado através do prisma da Ciência da Informação.

Figura 1: Distribuição do Zika Vírus – Américas.



Fonte: Organização Mundial de Saúde Apud Dias(2016).

73 Pesquisa financiado pelo CNPq. Processo 427308/2016-0

O Zika Vírus tem como vetores de transmissão os mosquitos denominados de *Aedes aegypti* e *Aedes albopictus*, o primeiro com abrangência apenas em regiões tropicais e subtropicais, o segundo podendo sobreviver e hibernar em regiões de clima mais frio (WHO, 2016).

O Zika Vírus tem como vetores de transmissão os mosquitos denominados de *Aedes aegypti* e *Aedes albopictus*, o primeiro com abrangência apenas em regiões tropicais e subtropicais, o segundo podendo sobreviver e hibernar em regiões de clima mais frio (WHO, 2016).

No Brasil, a eclosão de casos de doenças hipoteticamente associadas ao Zika Vírus, de forma mais notória a microcefalia, tem como epicentro os estados de Pernambuco e da Paraíba (DIAS, 2016).

A disseminação de informações sobre o Zika Vírus, tanto nos círculos acadêmicos e científicos formais, como nos informais na mídia de massa através das plataformas de rede social, apresentam inúmeras oportunidades de investigação para o pesquisador em Ciência da Informação. É nesta vertente que esta pesquisa está estruturada, através da análise de dados através da Análise de Redes Sociais (ARS) e de ferramentas de visualização que possibilitam a compreensão da disseminação de informações relacionadas com a epidemia de Zika Vírus, contribuindo com os resultados obtidos para o campo científico e para a sociedade, visto a urgência e relevância da temática (DIAS, 2016).

2 Objetivos

A pesquisa em tela investiga de que maneira a Ciência da Informação pode fornecer suporte para entendimento de questões que emergem a partir da disseminação de informações acerca do Zika vírus postadas na rede social Twitter. O objetivo principal é investigar a disseminação de informações através de canais formais e informais como redutora de incertezas de uma população imersa em epidemia de Zika Vírus. Os objetivos específicos são: Mapear o fluxo informacional sobre o Zika Vírus no âmbito do Twitter; Identificar as especificidades da disseminação das informações acerca do Zika Vírus no Twitter; Desenvolver software para coletar dados sobre o tema Zika Vírus no Twitter; Produzir visualizações de dados a partir dos dados coletados para gerar novos padrões associativos; Gerar métricas associadas aos dados coletados no Twitter (DIAS, 2016).

3 Procedimentos metodológicos

A pesquisa ora em andamento, configura-se com relação aos procedimentos técnicos como bibliográfica, com relação aos objetivos é exploratória e possui natureza quantitativa e qualitativa (GIL, 2008).

O corpus estudado constituiu-se de mensagens disseminadas (tweets⁷⁴) através da rede social Twitter com conteúdo relacionado ao Zika Vírus. Diversos padrões de grafias associados ao Zika Vírus foram utilizados nas capturas da mensagem. Esta estratégia teve como por objetivo possibilitar a coleta do maior número de mensagens possível relacionadas com o Zika Vírus.

Para a captura dos tweets foi elaborado um script⁷⁵ na linguagem de programação Ruby⁷⁶. Esta linguagem foi escolhida por permitir o desenvolvimento rápido de aplicações através de

74 Um tweet é considerado uma postagem na rede social Twitter.

75 Script Disponível a partir do link <http://wrco.ccsa.ufpb.br/zika.rb>

76 <https://www.ruby-lang.org/pt/>

acesso simplificado a Application Programming Interface (API) do Twitter e por ser de fácil aprendizado. Os tweets foram obtidos no padrão JavaScript Object Notation⁷⁷(JSON).

O script para a captura dos tweets foi implantado em um servidor Debian Linux⁷⁸ e capturou mensagens postadas no período compreendido entre outubro de 2017 até março de 2018, integralizando um total de 6 (seis) meses. Os tweets capturados foram armazenados em banco de dados PostgreSQL através do tipo de dados JSON. O script capturou durante o seu período de execução um total de 2.907.720 tweets, das quais mais de 82% são mensagens escritas nas línguas portuguesa, inglesa e espanhola. As consultas ao corpus foram feitas através da Structured Query Language (SQL). Indicamos como obstáculo para a captura dos tweets as eventuais quedas do servidor Debian Linux em virtude de interrupções no fornecimento de energia elétrica na Universidade Federal da Paraíba nos finais de semanas e as interrupções no acesso à internet. Esta situação foi minimizada com ajustes na configuração de parâmetros entre o sistema operacional Debian Linux e a aplicação Ruby para que esta segunda tentasse se recuperar após alguma condição de instabilidade do sistema.

Os tweets capturados foram processados mediante o uso da Linguagem SQL, no ambiente provido pelo Sistema Gerenciador de Bancos de Dados (SGBD) PostgreSQL⁷⁹, com a função de exportação dos resultados em formato CSV. Os arquivos com os dados eram então, posteriormente, visualizados ou tratados no SublimeText⁸⁰ e/ou OpenOffice⁸¹.

De modo a facilitar a análise dos tweets, foram criados dois arquivos em formato texto. Um arquivo constituído de tweets na língua portuguesa e outro composto por tweets na língua inglesa. Devido à estrutura JSON do tweet conter, em caso de retweet, toda a estrutura de dados do tweet original, o comando SQL buscava por combinações distintas de data/hora e texto, conforme exposto no código da Figura 2. Com isto, garante-se que os tweets originais apareçam apenas uma vez, mas também permite que no caso de dois tweets onde o texto seja o mesmo, por ter os dados de data/hora distintos, ambos sejam recuperados.

Figura 2: Comando SQL para recuperar data/hora de criação e texto do tweet.

```
select
  distinct(
    dados->'created_at', texto)
  from tweets
  where dados->>'lang' = 'en' or
        dados->>'lang' = 'pt');
```

Fonte: Desenvolvimento nosso.

Para recuperar as hashtags de cada tweet, foi utilizado o comando SQL aqui exibido na Figura 3.

77 <https://www.json.org/>

78 <https://www.debian.org/>

79 <https://www.postgresql.org/>

80 <https://www.sublimetext.com/3>

81 <https://www.openoffice.org>

Figura 3: Comando SQL para recuperar hashtags dos tweets.

```
select dados->'lang',
       upper(jsonb_array_elements(
         dados->'entities'-
         >'hashtags')->>'text'),
       count(*)
  from tweets
 where
   dados->'entities'-
   >'hashtags' != ''''' and
   dados->'entities'-
   >'hashtags' != '[""]'
  group by 1, 2;
```

Fonte: Desenvolvimento nosso.

4 Resultados preliminares

As análises iniciais dos arquivos de tweets revelaram alguns fatos curiosos. As postagens de tweets na língua portuguesa continham pouca relação com a questão da Zika e assuntos que remetesse ao tema saúde, a vasta maioria das postagens estão relacionadas com o uso do termo “Zika” em questões coloquiais do dia-a-dia das pessoas. O Quadro 1 ilustra esta situação.

Quadro 1: Amostra dos conteúdos das mensagens dos tweets capturados.

Mensagem do tweet
Hoje a sessão foi zika
@saladeredacao Caiu por causa do Baldasso, homem Zika
Quando eu casar vou entrar na igreja com um Nike Zika no pé, nada de sapato social...
essa semana que vou ficar sozinho em casa vai ser mais que zika
será q saiu minha zika do fifa
meu gatinho curioso vai dar dengue de tão parado
Hoje foi zika
Meu treinador é Deus, que me botou pra jogar olhou pro campo e disse zica vai lá
Minha vida sexual tá mais parada que Água pra dengue
Hoje a sessão foi zika

Fonte: Dados de pesquisa.

O arquivo texto com tweets em língua portuguesa também foi investigado para se identificar quais as tags mais frequentes no contexto das postagens associadas ao tema Zika. O Quadro 2, apresentado a seguir ilustra os resultados encontrados.

Além das informações relacionadas com o sentido das mensagens disseminadas no Twitter e das hashtags mais frequentes encontradas no corpus, os pesquisadores associados à pesquisa estão investigando parâmetros relacionados com a localização geográfica das mensagens e a aderência dos tweets a assuntos que remetam a questões de saúde.

Quadro 2: Hashtags mais frequentes nos tweets em Língua Portuguesa.

Hashtag	Qtd
MALHAÇÃO	3023
FLAMENGO	1860
DENGUE	1840
ZIKA	1787
ZICO	1770
ZICAÇO	1759
ZICÃO	1751
BBB18	1666
VIVAZICO	1612
TREZEDIASLONGEDOSOL	863

Fonte: Dados de pesquisa.

5 Considerações parciais

Até o estágio atual do desenvolvimento da pesquisa os objetivos específicos elencados foram parcialmente atingidos. Em uma primeira instância é possível mencionar que os usuários do Twitter que postam mensagens com o termo “zica” e strings equivalentes, não estão na maioria das vezes fazendo referência às questões de saúde associadas com a epidemia de Zika Vírus.

O termo zica (zika) tem uso em uma grande variedade de contextos que transcendem o domínio específico da área da saúde. Este uso mostrou-se predominante. É preciso empreender estudos mais aprofundados para entender os motivos que levam a população de Língua Portuguesa (Brasil) a não usar o Twitter como uma ferramenta que possibilite a disseminação de mensagens acerca de um problema de saúde pública, ferramenta esta que poderia ser útil na obtenção de informações sobre o combate ao mosquito vetor do Zika Vírus, bem como no tratamento das enfermidades associadas a este vírus.

Referências

- DIAS, G. A. **A Ciência da Informação e a disseminação de informações associadas a epidemia de Zika vírus: uma investigação baseada na Análise de Redes Sociais.** João Pessoa. Universidade Federal da Paraíba. 2016. Projeto de Pesquisa. Chamada Universal – MCTI/CNPq N. 01/2016
- DUARTE, Z.; FARIAS, L. (Orgs). **A medicina na era da informação.** Salvador: EDUFBA, 2009.
- GALVAO, M. C. B.; RICARTE, I. L. M. **Prontuário do Paciente.** Rio de Janeiro: Guanabara Koogan, 2012.
- GIL, A. C. **Como elaborar projetos de pesquisa.** 4. ed. São Paulo: Atlas, 2008.
- GOMEZ, M. N. G. O Domínio das Informações em Saúde. In: PINTO, V. B.; CAMPOS, H. H. (Org). **Diálogos Paradigmáticos sobre informação para a área da saúde.** Fortaleza: Edições UFC, 2013.

PINTO, V. B.; CAMPOS, H. H. (Org). **Diálogos Paradigmáticos sobre informação para a área da saúde**. Fortaleza: Edições UFC, 2013.

WHO. (2016). **Zika virus disease: Questions and answers**. Disponível em: <<http://www.who.int/features/qa/zika/en/>>. Acesso em: 03 fev. 2018.