

Ricardo César Gonçalves Sant'Ana

Moisés Lima Dutra

Guilherme Ataíde Dias

Organizadores

WIDaT 2018

II WORKSHOP DE INFORMAÇÃO,
DADOS E TECNOLOGIA

ANAIS
WIDaT 2018

Organização do WIDaT 2018

- **Organização Geral:**

Guilherme Ataíde Dias (PPGCI-UFPB) - Coordenador geral do evento
Moisés Lima Dutra (PPGCIN-UFSC) - Vice-coordenador

- **Coordenador da Comissão Científica:**

Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)

- **Comissão científica**

Adilson Luiz Pinto (PPGCIN-UFSC)
Ana Alice Baptista (Universidade do Minho, Portugal)
Ana Carolina Simionato (PPGCI-UFSCar)
Angela Maria Grossi de Carvalho (PPGCI-UNESP)
Bernardina Maria Juvenal Freire de Oliveira (PPGCI-UFPB)
Cristian Berrío-Zapata (PPGCI-UFPA)
Dalton Lopes Martins (FCI-UnB)
Denysson Axel Ribeiro Mota (PPGB-UFCA)
Douglas Dyllon Jeronimo de Macedo (PPGCIN-UFSC)
Ed Porto Bezerra (PPGI-UFPB)
Edgar Bisset Alvarez (PPGCIN-UFSC)
Edna Gusmão de Goés Brennand (MPGOA-UFPB)
Edna Gomes Pinheiro (DCI-UFPB)
Elaine Parra Affonso (FATEC-SP)
Elvis Fusco (UNIVEM-Marília)
Enrique Muriel Torrado (PPGCIN-UFSC)
Evandro de Barros Costa (IC-UFAL)
Fábio Paraguaçu (IC-UFAL)
Fernando de Assis Rodrigues (PPGCI-UNESP)
Gustavo Medeiros de Araújo (PPGCIN-UFSC)
Henry Pôncio Cruz de Oliveira (PPGCI-UFPB)
Joana Coeli Ribeiro Garcia (PPGCI-UFPB)
José Eduardo Santarém Segundo (USP-FFCLRP)
Leonardo Castro Botega (UNIVEM-Marília)
Luana Farias Sales Marques (PPGCI-IBICT-UFRJ)
Marckson Roberto Ferreira de Sousa (PPGCI-UFPB)
Luís Fernando Sayão (CNEN)
Marcelo Morandini (EACH-USP)
Márcio Matias (PPGCIN-UFSC)
Marcos Mucheroni (CBD-USP)
Marynice de Medeiros Matos Autran (PPGCI-UFPB)

Maurício Barcellos Almeida (PPGGOC-UFMG)
Moisés Lima Dutra (PPGCIN-UFSC)
Plácida Leopoldina V. da Costa Santos (PPGCI-UNESP)
Pedro Luiz Pizzigatti Corrêa (POLI-USP)
Renata Baracho (PPGGOC-UFMG)
Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)
Robson Rodrigues Lemos (UFSC-Araranguá)
Rogério Ramalho (PPGCI-UFSCar)
Ryan Ribeiro de Azevedo (UFRPE-UAG)
Sandra de Albuquerque Siebra (PPGCI-UFPE)
Sandro Rautenberg (DECOMP-UNICENTRO)
Silvana Aparecida Borsetti G. Vidotti (PPGCI-UNESP)
Virginia Bentes Pinto (PPGCI-UFC)
Wagner Junqueira de Araújo (PPGCI-UFPB)
Zaira Regina Zafalon (PPGCI-UFSCar)

- **Coordenador do Cerimonial:**

André Luiz Dias de França (PPGCI-UFPB)

- **Coordenador da Equipe Técnica Local:**

Laerte Pereira da Silva Júnior (CCHLA-UFPB)

- **Equipe Técnica Local:**

Adriana Alves Rodrigues (PPGCI-UFPB)
Antonio Felipe dos Santos (MPGOA-UFPB)
Débora Gomes de Araújo (PPGCI-UFPB)
Pedro Augusto de Lima Barroso (PPGCI-UFPB)
Pollianna Marys de Souza e Silva (PPGCI-UFPB)
Renata Lemos dos Anjos (PPGCI-UFPB)

A DISSEMINAÇÃO DE INFORMAÇÕES ASSOCIADAS A EPIDEMIA DO ZIKA VÍRUS: Uma pesquisa a partir de dados coletados no *Twitter*

THE DISSEMINATION OF INFORMATION ASSOCIATED WITH THE ZIKA VIRUS EPIDEMIC:
A research based on data collected from the *Twitter*

Guilherme Ataíde Dias¹, Denysson Axel Ribeiro Mota², André Luiz Dias de França³, Eduardo de Santana Medeiros Alexandre⁴

(1) Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, guilhermeataide@ccsa.ufpb.br.

(2) Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, denysson.mota@ufca.edu.br.

(3) Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, andreluizjpb@gmail.com.

(4) Instituto Federal da Paraíba, eduardo.ufpb@gmail.com.

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.

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.

Keywords: *Twitter*; Data dissemination; Zika virus.

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 (*tweets*) através da rede social *Twitter* com conteúdo relacionado ao Zika Vírus. Diversos padrões de grafias associados ao Zika Vírus foram utilizados na captura das mensagens. 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 *script* na linguagem de programação *Ruby*. Esta linguagem foi escolhida por permitir o desenvolvimento rápido de aplicações através de acesso simplificado a *Application Programming Interface* (API) do *Twitter* e por ser de fácil aprendizado. Os *tweets* foram obtidos no padrão *JavaScript Object Notation* (JSON).

O *script* para a captura dos *tweets* foi implantado em um servidor *Debian Linux* 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 este segundo 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) *PostgreSQL*, com a função de exportação dos resultados em formato CSV. Os arquivos com os dados eram então, posteriormente, visualizados ou tratados no *SublimeText* e/ou *OpenOffice*.

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.

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
festinha ficou zika ta, curti
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 zika 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.

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
TREZEDIAS- LONGEDOSOL	863

Fonte: Dados de pesquisa

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.

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.