



**UNIVERSIDADE FEDERAL DA FRONTEIRA SUL  
CAMPUS DE CHAPECÓ  
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**JOD FEDLET PIERRE**

**UTILIZAÇÃO DE MODELAGEM DE TÓPICOS PARA IDENTIFICAR OS  
ASSUNTOS MAIS DISCUTIDOS SOBRE O HAITI NO TWITTER**

**CHAPECÓ  
2022**

**JOD FEDLET PIERRE**

**UTILIZAÇÃO DE MODELAGEM DE TÓPICOS PARA IDENTIFICAR OS  
ASSUNTOS MAIS DISCUTIDOS SOBRE O HAITI NO TWITTER**

Trabalho de conclusão de curso apresentado como requisito para obtenção do grau de Bacharel em Ciência da Computação da Universidade Federal da Fronteira Sul.  
Orientador: Dr. Denio Duarte

**CHAPECÓ**  
**2022**

Pierre, Jod Fedlet

Utilização de modelagem de tópicos para identificar os assuntos mais discutidos sobre o Haiti no Twitter / Jod Fedlet Pierre. – 2022.

61 f.: il.

Orientador: Dr. Denio Duarte.

Trabalho de conclusão de curso (graduação) – Universidade Federal da Fronteira Sul, curso de Ciência da Computação, Chapecó, SC, 2022.

1. Modelagem de tópicos. 2. Haiti. 3. Twitter. 4. LDA. I. Duarte, Dr. Denio, orientador. II. Universidade Federal da Fronteira Sul. III. Título.

**JOD FEDLET PIERRE**

**UTILIZAÇÃO DE MODELAGEM DE TÓPICOS PARA IDENTIFICAR OS  
ASSUNTOS MAIS DISCUTIDOS SOBRE O HAITI NO TWITTER**

Trabalho de conclusão de curso apresentado como requisito para obtenção do grau de Bacharel em Ciência da Computação da Universidade Federal da Fronteira Sul.

Orientador: Dr. Denio Duarte

Aprovado em: \_\_/\_\_/\_\_\_\_.

BANCA AVALIADORA

---

Dr. Denio Duarte – UFFS

---

Prof. Dr. Felipe Grando – UFFS

---

Prof. Me. Geomar Schreiner – UFFS

## AGRADECIMENTOS

Primeiramente, quero agradecer a Deus por me conceder vida, saúde e capacidade para chegar até o ponto final dessa trajetória.

Agradeço aos meus irmãos: Felix Pierre Delus, Isemane Pierre Delus, Ovanel Pierre, Renante Pierre, Michelet Pierre, Renide Pierre e Dukenson Pierre que sempre me apoiaram e me incentivaram para que pudesse chegar até aqui.

Aos colegas (novos membros da família) Eddy Celestin e Abdias Revolte, com quem convivi desde o início do caminho até o presente momento, pelo companheirismo e apoio que me permitiram crescer como pessoa e formando. Inclusive agradeço à Sarah André por ter me apoiado durante vários momentos de desânimo.

À Universidade Federal da Fronteira Sul pela oportunidade que oferece aos imigrantes para ingressar numa universidade gratuita e de boa qualidade, mediante o programa Pro-Haiti.

Aos responsáveis do Pro-Haiti (Programa de Acesso à Educação Superior da UFFS para Estudantes Haitianos) pela luta que enfrentam pela manutenção desse projeto.

Ao Setor de Assuntos Estudantis do Campus Chapecó (ASSAE-CH) que tem me apoiado financeiramente ao longo do caminho.

Agradeço a todos os professores pelas correções e ensinamentos que me permitiram apresentar um melhor desempenho no meu processo de formação profissional ao longo do curso.

Agradeço ao meu orientador Dr. Denio Duarte pela sugestão do tema deste trabalho, ensinamentos, paciência, revisões e orientações. Também aos membros das bancas que contribuíram para o aprimoramento deste trabalho: Dr. Felipe Grando e Me. Geomar Schreiner.

Aos meus colegas de turma, por compartilharem comigo tantos momentos de descobertas e aprendizado e por todo o companheirismo ao longo deste percurso.

A todos que contribuíram de uma forma ou de outra para que conseguisse chegar até aqui.

*“Un gagnant est un rêveur qui n’abandonne jamais.”, Nelson Mandela*

## RESUMO

Com o crescimento do acesso a internet da população mundial, grandes volumes de dados, principalmente no formato de textos, estão sendo compartilhados diariamente. Seja por meio de redes sociais, blogs ou fóruns, esses dados (ou coleção de documentos) circulam na web de forma contínua pelos usuários. O Twitter é um dos meios mais utilizados pelos internautas para disseminação de informação através de postagens. Essas postagens, chamadas *tweets*, podem ser analisadas, interpretadas, compreendidas ou classificadas em grupos. Por serem compartilhados em quantidade enorme, foi-se necessário a utilização de técnicas de modelagem de tópicos, conjunto de algoritmos que ajudam a agrupar coleção de documentos em tópicos. Posto isso, neste trabalho, foi utilizado a Latent Dirichlet Allocation (LDA), como abordagem de modelagem de tópicos, em *tweets* discutidos sobre o Haiti em língua portuguesa durante janeiro de 2018 a maio de 2022 para descobrir os assuntos mais recorrentes. A coleção do experimento foi dividida em 53 meses para um melhor proveito dos resultados. Os resultados obtidos são apresentados por meio de tabelas e gráficos e são discutidos mensalmente a cada ano. Através desses resultados, é possível constatar os assuntos mais discutidos durante os períodos definidos.

Palavras-chave: Modelagem de tópicos. Haiti. Twitter. LDA.

## **ABSTRACT**

With the growth of internet access by the world's population, large volumes of data, mainly in text format, are being shared daily. Whether through social networks, blogs or forums, this data (or collection of documents) is continuously circulated on the web by users. Twitter is one of the most used means by Internet users to disseminate information through posts. These posts, called tweets, can be analyzed, interpreted, understood or classified into groups. Because they are shared in huge numbers, it was necessary to use topic modeling techniques, a set of algorithms that help group collections of documents into topics. That said, in this work, Latent Dirichlet Allocation (LDA), as a topic modeling approach, was used on tweets discussed about Haiti in Portuguese language during January 2018 to May 2022 to discover the most recurring topics. The experiment collection was divided into 53 months for a better use of the results. The results obtained are presented through tables and graphs and are discussed monthly each year. Through these results, it is possible to see the most discussed subjects during the defined periods.

Keywords: Topic modeling. Twitter. Haiti. LDA.

## RÉSUMÉ

Avec le développement de l'accès à l'internet par la population mondiale, de grands volumes de données, principalement sous forme de texte, sont partagés quotidiennement. Que ce soit par le biais des réseaux sociaux, des blogs ou des forums, ces données (ou collections de documents) circulent en permanence sur le web par les utilisateurs. Twitter est l'un des moyens les plus utilisés par les internautes pour diffuser des informations par le biais de messages. Ces messages, appelés *tweets*, peuvent être analysés, interprétés, compris ou classés en groupes. Comme ils sont partagés en grandes quantités, il était nécessaire d'utiliser des techniques de modélisation des sujets, un ensemble d'algorithmes qui aident à regrouper des collections de documents en sujets. Cela dit, dans ce travail, l'allocation de Dirichlet latente (LDA), en tant qu'approche de modélisation des sujets, a été utilisée sur les *tweets* discutés sur Haïti en langue portugaise de janvier 2018 à mai 2022 pour découvrir les sujets les plus récurrents. La collection d'expériences a été divisée en 53 mois pour une meilleure utilisation des résultats. Les résultats obtenus sont présentés sous forme de tableaux et de graphiques et font l'objet d'une discussion mensuelle chaque année. Grâce à ces résultats, il est possible de voir les sujets les plus discutés pendant les périodes définies.

Mots-clé: Modélisation de thèmes. Twitter. Haïti. LDA.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Posição geográfica da República do Haiti. . . . .	15
Figura 2 – Exemplo de perfil no Twitter. . . . .	19
Figura 3 – Exemplo de <i>tweet</i> publicado sobre o Haiti. . . . .	19
Figura 4 – Representação gráfica do modelo LDA. . . . .	22
Figura 5 – Exemplo de <i>tweets</i> extraídos na base de dados do Twitter. . . . .	24
Figura 6 – Exemplo de <i>tweets</i> antes e depois do pré-processamento dos dados. . . . .	25
Figura 7 – Tópicos obtidos a partir dos <i>tweets</i> extraídos do dia 05 de Agosto de 2016 . . . . .	26
Figura 8 – Nuvem de palavras obtidas em um dos dez tópicos. . . . .	27
Figura 9 – Exemplo de rotulação usando notícias do <i>Google</i> . . . . .	33
Figura 10 – Frequência dos tópicos predominantes de cada mês de 2018. . . . .	35
Figura 11 – Frequência dos tópicos predominantes de cada mês de 2019. . . . .	37
Figura 12 – Frequência dos tópicos predominantes de cada mês de 2020. . . . .	39
Figura 13 – Frequência dos tópicos predominantes de cada mês de 2021. . . . .	41
Figura 14 – Frequência dos tópicos predominantes dos cinco meses de 2022. . . . .	42

## LISTA DE TABELAS

Tabela 1 – Três tópicos com as suas top-5 palavras e probabilidades. . . . .	21
Tabela 2 – Exemplo de <i>tweet</i> antes e depois do pré-processamento dos dados. . . . .	31
Tabela 3 – Repartição anual dos <i>tweets</i> antes e depois do pré-processamento. . . . .	31
Tabela 4 – Tópico predominante de cada mês de 2018. . . . .	34
Tabela 5 – Tópico predominante de cada mês de 2019. . . . .	36
Tabela 6 – Tópico predominante de cada mês de 2020. . . . .	38
Tabela 7 – Tópico predominante de cada mês de 2021. . . . .	40
Tabela 8 – Tópico predominante de cada mês de 2022. . . . .	41
Tabela 9 – Tópicos de janeiro de 2018. . . . .	48
Tabela 10 – Tópicos de fevereiro de 2018. . . . .	48
Tabela 11 – Tópicos de março de 2018. . . . .	48
Tabela 12 – Tópicos de abril de 2018. . . . .	48
Tabela 13 – Tópicos de maio de 2018. . . . .	49
Tabela 14 – Tópicos de junho de 2018. . . . .	49
Tabela 15 – Tópicos de julho de 2018. . . . .	49
Tabela 16 – Tópicos de agosto de 2018. . . . .	49
Tabela 17 – Tópicos de setembro de 2018. . . . .	49
Tabela 18 – Tópicos de outubro de 2018. . . . .	49
Tabela 19 – Tópicos de novembro de 2018. . . . .	49
Tabela 20 – Tópicos de dezembro de 2018. . . . .	50
Tabela 21 – Tópicos de janeiro de 2019. . . . .	51
Tabela 22 – Tópicos de fevereiro de 2019. . . . .	51
Tabela 23 – Tópicos de março de 2019. . . . .	51
Tabela 24 – Tópicos de abril de 2019. . . . .	51
Tabela 25 – Tópicos de maio de 2019. . . . .	52
Tabela 26 – Tópicos de junho de 2019. . . . .	52
Tabela 27 – Tópicos de julho de 2019. . . . .	52
Tabela 28 – Tópicos de agosto de 2019. . . . .	52
Tabela 29 – Tópicos de setembro de 2019. . . . .	52
Tabela 30 – Tópicos de outubro de 2019. . . . .	52
Tabela 31 – Tópicos de novembro de 2019. . . . .	52
Tabela 32 – Tópicos de dezembro de 2019. . . . .	53
Tabela 33 – Tópicos de janeiro de 2020. . . . .	54
Tabela 34 – Tópicos de fevereiro de 2020. . . . .	54
Tabela 35 – Tópicos de março de 2020. . . . .	54
Tabela 36 – Tópicos de abril de 2020. . . . .	54
Tabela 37 – Tópicos de maio de 2020. . . . .	55

Tabela 38 – Tópicos de junho de 2020. . . . .	55
Tabela 39 – Tópicos de julho de 2020. . . . .	55
Tabela 40 – Tópicos de agosto de 2020. . . . .	55
Tabela 41 – Tópicos de setembro de 2020. . . . .	55
Tabela 42 – Tópicos de outubro de 2020. . . . .	55
Tabela 43 – Tópicos de novembro de 2020. . . . .	55
Tabela 44 – Tópicos de dezembro de 2020. . . . .	56
Tabela 45 – Tópicos de janeiro de 2021. . . . .	57
Tabela 46 – Tópicos de fevereiro de 2021. . . . .	57
Tabela 47 – Tópicos de março de 2021. . . . .	57
Tabela 48 – Tópicos de abril de 2021. . . . .	57
Tabela 49 – Tópicos de maio de 2021. . . . .	58
Tabela 50 – Tópicos de junho de 2021. . . . .	58
Tabela 51 – Tópicos de julho de 2021. . . . .	58
Tabela 52 – Tópicos de agosto de 2021. . . . .	58
Tabela 53 – Tópicos de setembro de 2021. . . . .	58
Tabela 54 – Tópicos de outubro de 2021. . . . .	58
Tabela 55 – Tópicos de novembro de 2021. . . . .	58
Tabela 56 – Tópicos de dezembro de 2021. . . . .	59
Tabela 57 – Tópicos de janeiro de 2022. . . . .	60
Tabela 58 – Tópicos de fevereiro de 2022. . . . .	60
Tabela 59 – Tópicos de março de 2022. . . . .	60
Tabela 60 – Tópicos de abril de 2022. . . . .	60
Tabela 61 – Tópicos de maio de 2022. . . . .	61

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>12</b>
1.1	ESTRUTURA DO TRABALHO . . . . .	14
<b>2</b>	<b>HAITI E TWITTER . . . . .</b>	<b>15</b>
2.1	HAITI . . . . .	15
<b>2.1.1</b>	<b>Breve Histórico . . . . .</b>	<b>15</b>
<b>2.1.2</b>	<b>Fatos sobre o Haiti . . . . .</b>	<b>16</b>
2.2	TWITTER . . . . .	18
<b>3</b>	<b>MODELAGEM DE TÓPICOS . . . . .</b>	<b>21</b>
<b>4</b>	<b>TRABALHOS RELACIONADOS . . . . .</b>	<b>24</b>
<b>5</b>	<b>METODOLOGIA . . . . .</b>	<b>29</b>
5.1	FERRAMENTAS . . . . .	29
5.2	COLETA DOS DADOS . . . . .	29
5.3	PRÉ-PROCESSAMENTO DOS DADOS . . . . .	30
5.4	APLICAÇÃO DO MODELO LDA . . . . .	31
<b>5.4.1</b>	<b>Treinamento do modelo e identificação do melhor número de tópicos . .</b>	<b>31</b>
<b>5.4.2</b>	<b>Extração dos tópicos e identificação dos mais frequentes em cada mês .</b>	<b>32</b>
5.5	ROTULAÇÃO MANUAL DOS TÓPICOS . . . . .	33
<b>6</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>34</b>
6.1	2018 . . . . .	34
6.2	2019 . . . . .	36
6.3	2020 . . . . .	37
6.4	2021 . . . . .	39
6.5	2022 . . . . .	41
<b>7</b>	<b>CONCLUSÃO . . . . .</b>	<b>44</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>45</b>
<b>A</b>	<b>APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2018 . . . . .</b>	<b>48</b>
<b>B</b>	<b>APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2019 . . . . .</b>	<b>51</b>
<b>C</b>	<b>APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2020 . . . . .</b>	<b>54</b>
<b>D</b>	<b>APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2021 . . . . .</b>	<b>57</b>
<b>E</b>	<b>APÊNDICE - TÓPICOS EXTRAÍDOS NOS CINCO MESES DO ANO DE 2022 . . . . .</b>	<b>60</b>

## 1 INTRODUÇÃO

A internet é uma grande rede de informação que está evoluindo constantemente. Por consequência, grandes volumes de dados estão sendo compartilhados na web entre usuários que não precisam estar necessariamente na mesma posição geográfica. Esses dados podem transmitir informações importantes referentes a diversos assuntos mundiais como, por exemplo, política, jogos e cultura. Geralmente, esses dados circulam na web no formato de textos por meio de redes sociais, blogs ou fóruns e são escritos de maneira livre.

O Twitter, lançado em 2006, é um exemplo de rede social muito utilizada pelos usuários a fim de disseminar conteúdos na internet a respeito de assuntos de interesse (TWITTER, 2020). É uma plataforma muito aberta que também proporciona acesso fácil à informação. Atualmente, a plataforma conta com 300 milhões de usuários que compartilham informações diariamente por meio de postagens. Essas postagens são chamadas de *tweets* e possuem um limite de até 280 caracteres (ROHR; SQUIRES; PETERS, 2022).

Essas postagens podem ser classificadas, organizadas, resumidas ou agrupadas por tópicos. O agrupamento dos dados ou coleção de documentos em tópicos, por serem dados compartilhados em quantidade enorme, requer muito tempo para a análise manual. Logo, uma abordagem para agrupar discussões por assunto é a modelagem de tópicos.

A modelagem de tópicos é um conjunto de algoritmos que visam analisar palavras em grande coleção de documentos a fim de descobrir temas ou assuntos principais. Geralmente, esses documentos são escritos de maneira livre, pois são desestruturados. Esses algoritmos (ou abordagens) também ajudam a organizar e/ou resumir esses dados (BLEI, 2012).

A modelagem de tópicos baseia-se na ideia de que os documentos são misturas de tópicos, em que um tópico é uma distribuição de probabilidade sobre palavras. Ou seja, os documentos podem exibir vários tópicos, no qual cada tópico é formado por um conjunto de palavras, sendo que cada palavra do tópico possui uma probabilidade que representa a importância da palavra no tópico. Da mesma maneira, uma coleção é um conjunto de documentos e que cada documento pode ser qualquer conteúdo textual, por exemplo, artigo ou publicação em rede social (STEYVERS; GRIFFITHS, 2007). Neste trabalho, usou-se a modelagem de tópicos em *tweets* discutidos sobre o Haiti.

Haiti é um país que tem passado por muitos problemas ou desastres ao longo dos tempos. Sejam naturais ou causados por ações humanas. Dentre os desastres naturais, vários terremotos atingiram o país, em que o mais devastador foi o de 12 de janeiro de 2010. Nesse tremor, foram cerca de 300 mil mortos, 1,5 milhão de pessoas feridas e 1,5 milhão de pessoas desabrigadas. Além disso, várias instituições importantes foram destruídas. Por exemplo, o Palácio Nacional onde o presidente trabalha (CHAUHAN; SHAH, 2021). O assassinato do presidente Jovenel Moïse no dia 07 de julho de 2021 (TROUILLARD, 2021) é um dos eventos causados por ações humanas.

Essas situações adversas pioram as crises sociopolítico-econômicas pré-existentes no

país. Como consequência, a emigração haitiana, que anteriormente ao tremor de 2010 visava países como Estados Unidos da América, Canadá e França, começou a visar os países da América Latina como novo destino, principalmente o Brasil. Outro fator que propiciou a escolha do Brasil como principal destino foi a liderança do Brasil na Missão das Nações Unidas para a Estabilização do Haiti de 2004 a 2017. Essa missão estreitou a relação existente entre os dois países (VERENHITACH et al., 2008). Além disso, o Brasil, por ter sido a sede da Copa do Mundo de 2014, incentivou a emigração haitiana em busca de novas oportunidades de trabalhos em construção civil.

Neste contexto, este trabalho objetiva-se explorar e analisar os principais assuntos discutidos sobre o Haiti na plataforma do Twitter, em língua portuguesa, usando LDA como técnica de modelagem de tópicos. A LDA é uma abordagem de modelagem de tópicos, considerada como a mais simples para extração de tópicos em coleção de documentos (DAVID M. BLEI ANDREW Y. NG, 2003).

Para alcançar os objetivos, foi realizada coleta dos dados na base do Twitter nas datas entre 01 de janeiro de 2018 a 31 de maio de 2022, com palavras-chave 'haiti', 'haitiano' e 'haitiana'. Obteve-se um total de 185.348 *tweets* e os mesmos passaram por um processo de pré-processamento, procedimento em que os dados indesejados foram removidos, por exemplo *links* e menções de perfis de usuários. Após o pré-processamento, o total de *tweets* reduziu para 150.174.

Em seguida, os dados pré-processados são repartidos em grupos mensais, 53 no total, e aplicados o modelo LDA. Usou-se métrica de coerência  $c_v$ , avaliação do modelo que permitiu encontrar o melhor número de tópicos de cada mês. Assim, considerou-se o número que se repete em mais meses para ser o número de tópicos a serem extraídos no estudo, sendo 5 esse número, denotado por  $K$ . Posteriormente, calculou-se a frequência dos tópicos obtidos em cada documento. Como também, realizou-se a rotulação do tópico mais frequente de cada mês. A rotulação é uma técnica que consiste em atribuir um tema ou assunto aos tópicos, baseado em conhecimentos empíricos da coleção ou pela análise das palavras mais importantes de cada tópico.

O uso de *tweets* permitiu identificar os diversos acontecimentos sobre o Haiti na visão dos usuários no período selecionado. A modelagem de tópicos facilitou o entendimento dos assuntos mais abordados no Twitter em língua portuguesa, por meio de agrupamento das palavras relacionadas dos documentos em tópicos. Percebeu-se, mediante discussão dos tópicos, que os resultados refletiram fatos reais relacionados ao Haiti.

As principais contribuições deste estudo podem ser apresentadas de forma social, política e econômica. Pode-se ajudar os leitores a pensarem nas atualidades sociopolítico-econômicas do Haiti, ver como os internautas lusófonos pensam, concebem ou tratam o país mediante a discussão dos tópicos. O Twitter, por ser um espaço de livre expressão, as pessoas o usam como tribunal para julgar. Assim como serve para passar frustrações e nervosismo ou para propor soluções aos diversos problemas que o país enfrenta.

## 1.1 ESTRUTURA DO TRABALHO

O restante deste trabalho está organizado da seguinte forma: os capítulos 2 e 3 apresentam a revisão bibliográfica, detalhando melhor os conceitos sobre Haiti, Twitter e modelagem de tópicos; o capítulo 4 apresenta os trabalhos relacionados, que serviram como guia para o estudo; no capítulo 5, é apresentada a metodologia empregada, contendo as etapas que ajudaram alcançar os objetivos do trabalho; na sequência, o capítulo 6 apresenta os resultados por meio de tabelas e gráficos e discussões desses resultados; por fim, o capítulo 7 apresenta a conclusão do estudo.

## 2 HAITI E TWITTER

Este capítulo apresenta alguns conceitos importantes para o entendimento deste trabalho. Inicialmente, é apresentado o país Haiti por ser foco dos *tweets* que serão estudados. Em seguida, é apresentada a rede social Twitter.

### 2.1 HAITI

A República do Haiti, comumente chamada Haiti, é um país localizado na América Central. O país é a parte oeste da ilha *Hispaniola*, partilhada com a República Dominicana (este último fica no lado leste). Em termos de população e de área, o Haiti é o terceiro maior país do Caribe (depois da Cuba e da República Dominicana), com uma superfície de 27.750 quilômetros quadrados e uma população ativa estimada de cerca de dez milhões de habitantes. A capital do Haiti é Porto Príncipe e os idiomas oficiais são francês e crioulo (FRANCISCO, 2022). A Figura 1 apresenta, em amarelo, a localização do Haiti em relação às Américas.



Figura 1 – Posição geográfica da República do Haiti.

Fonte: (FRANCISCO, 2022)

#### 2.1.1 Breve Histórico

Segundo (RTS, 2011), a ilha *Hispaniola*, habitada por populações indígenas, foi descoberta em 1492 por Cristovão Colombo. Essas populações foram massacradas pelo trabalho forçado (mineração de ouro) dos espanhóis. Para substituir essa força de trabalho, os colonos recorreram aos escravos africanos. Juntamente com os mulatos, os escravos africanos são os principais ancestrais da grande maioria dos haitianos.

Em meados do século XVI, os franceses se estabeleceram nas terras abandonadas pelos espanhóis, apesar dos esforços concentrados desses últimos para repeli-los. Os franceses,



uma vez instalados na ilha, também recorreram aos escravos africanos mas, para trabalhar nas plantações de café e açúcar. Em 1697, por meio do Tratado de Ryswick, os espanhóis cederam e reconheceram a soberania francesa sobre a parte ocidental da ilha onde é fundada a capital, Porto Príncipe, em 1749. Dentre todas as colônias do Novo Mundo, a chamada *La Saint-Domingue française* tornou-se a mais lucrativa, até a frente dos Estados Unidos (RTS, 2011).

Após várias décadas de trabalho forçado dos escravos negros, em 1791 iniciou uma rebelião de negros unidos com mulatas contra o regime escravista, principalmente no norte do território. Essa rebelião foi liderada principalmente por Toussaint Louverture, Jean-Jacques Dessalines, Henri Christophe e Alexandre Pétion. O evento foi seguido por conflitos civis e militares, muitas vezes ligados à agitação e ao fluxo político na França. O fim dessa revolta foi marcado pela capitulação do exército francês em 1803. Alguns meses depois, na cidade de Gonaïves, foi proclamada a independência do território, tornando o Haiti a primeira república negra livre do mundo e a segunda nação do continente americano, precedida apenas dos Estados Unidos. Essa revolta que conduziu até a independência do Haiti serviu como exemplo de superação de colônias e abriu os olhos de outros escravos do mundo (HURBON, 2018).

### **2.1.2 Fatos sobre o Haiti**

O Haiti é um dos países com mais vítimas por terremotos, por ser um país muito pobre e que não possui uma boa infraestrutura urbana. A falta de recursos, a precariedade das construções, as aglomerações urbanas e as fragilidades do Estado justificam os danos dos tremores (BBC, 2017). Além do mais, o país está localizado em uma área de elevada instabilidade tectônica. Além da colisão de diversas placas, duas extensas zonas de falha se estendem que são a placa norte-americana ao norte e a placa caribenha no sul do país, e a sua movimentação pode ser a causa de intensos tremores de terra. Como consequências, infraestrutura urbana, casas e edifícios assim como a perda humana, inúmeros feridos e pessoas desabrigadas são algumas das consequências desses tremores. Contudo, até hoje, o país está sofrendo as consequências de terremotos (GUITARRARA, 2022).

Dentre os terremotos que aconteceram no Haiti, aquele de 12 de janeiro de 2010 foi o mais devastador e aquele de 14 de agosto de 2021 é o mais recente. O de 12 de janeiro de 2010 ocorreu às 16 horas e 53 minutos e 10 segundos, horário local do Haiti. Considerou-se como o mais destrutivo e teve uma magnitude de entre 7,0 a 7,3 da escala de Richter. O epicentro do tremor foi localizado a aproximadamente a 25,3 km de Porto Príncipe, capital do país. Já que o seu hipocentro foi localizado a uma profundidade de 10 km (GUITARRARA, 2022).

Nesse terremoto, houve muitas mortes, vítimas, ferimentos e pessoas amputadas ou com perturbações psicológicas. Conforme os dados divulgados no dia 09 de fevereiro de 2010 pela Ministra das Comunicações, Marie-Laurence Jocelyn Lassegue, foram cerca de 300 mil mortos, 1,5 milhão de feridos e 1,5 milhão de pessoas desabrigadas. Além disso, muitas instituições importantes do país foram destruídas tais como o palácio nacional onde o presidente exerce o

seu trabalho e a Catedral da Nossa Senhora de Porto Príncipe (CHARLES, 2020).

Além das catástrofes naturais, o país tem enfrentado outras situações adversas e deploráveis como por exemplo, má gestão governamental, guerras civis e crises sociopolítico-econômicas. No entanto, essas situações fazem com que o povo reclame e às vezes, as reclamações acabam sendo feitas com violência. Apesar de mais de dois séculos já terem sido sucedidos a independência (1804), o país passa por essas situações difíceis e requer a cooperação da comunidade internacional. Os autores (VERENHITACH et al., 2008) relatam que essas crises proveem desses dois séculos de história independente, nos quais o Haiti sofreu 34 Golpes de Estado e promulgou 23 constituições, em um longo e constante ciclo de governos ditatoriais que resultou em um processo democrático incipiente, estabelecido sobre bases frágeis e contra a vontade da elite local.

Em vista disso, em 2004, iniciou-se a intervenção da comunidade internacional, em que foi constituída a Missão das Nações Unidas para a Estabilização do Haiti (MINUSTAH). A missão foi composta por 6700 militares e uma força policial internacional de 1622 homens. A missão tratava de impor o cumprimento da paz e não uma missão de manutenção da paz. Essa missão, liderada pelo Brasil, foi a primeira da Organização das Nações Unidas (ONU) sob a liderança de países latino-americanos (VERENHITACH et al., 2008).

Um dos danos mais recentes causados pela instabilidade política do país foi a morte do presidente, Jovenel Moïse. Moïse, nascido no dia 26 de 1968 na comuna Trou du Nord / Nordeste do Haiti, presidente da República do Haiti, eleito em 2016 e tomou posse no dia 07 de fevereiro de 2017, foi assassinado na madrugada entre o dia 06 e 07 de julho de 2021 (por volta de uma hora em sua residência privada em Pétienville), por um grupo armado não identificado. Nesse assassinato, a primeira dama, Martine Marie Etienne Moïse, foi ferida conforme a declaração do chefe do governo haitiano (TROUILLARD, 2021).

Conforme (BIASSETE, 2021), horas depois, essa morte deixou o país em desordem, o afundou novamente em uma crise após vários meses de crises institucionais pendentes. O parlamento não era mais funcional há um bom tempo por falta de organização de eleições legislativas a tempo. A tragédia deixou a população em uma grande preocupação, pois o país ficou sem presidente. Posteriormente à tragédia, a República Dominicana, país vizinho, ordenou o fechamento imediato de sua fronteira com o Haiti para evitar que os suspeitos tivessem deixado o país (Haiti).

Anteriormente ao terremoto de 2010, a emigração haitiana teve como destino a França, os Estados Unidos e o Canadá. Entretanto, não era de forma tão massiva como está sendo feita atualmente. O terremoto acabou piorando as crises econômicas, além dos conflitos políticos existentes no país. Assim, o povo começou a olhar outros destinos em busca de uma vida melhor. Seja por condições deficientes econômicas, sociais, educativas ou questões de segurança. São vários os motivos que levam os haitianos a deixarem o seu país de forma excessiva para a América Central, principalmente para o Brasil.

Em 2014, o Brasil preparava-se para sediar a Copa do Mundo, por consequência, a

construção civil estava em ampla extensão. Os haitianos olhavam para o Brasil com a esperança de encontrar boas oportunidades de trabalho e melhoria de vida. Outros fatores impulsivos de eles terem escolhido o Brasil como destino foi que a Missão das Nações Unidas para a Estabilização do Haiti, liderada pelo Brasil, estreitou a relação dos dois países (SOUZA, 2022). Além disso, a seleção brasileira de futebol possui uma grande torcida no Haiti e virou a paixão de muitos há muito tempo, como se fosse a seleção dos próprios haitianos.

## 2.2 TWITTER

Em março de 2006 foi lançado o Twitter, uma das redes sociais mais populares da atualidade (TWITTER, 2020). Conhecido também como um microblog, o Twitter permite que usuários compartilhem perguntas, sentimentos, ideias e opiniões referentes a vários assuntos, como política, saúde, economia, sociedade, música, cinema, celebridades e jogos por meio de *tweets*. Um *tweet* é uma publicação realizada na plataforma que, geralmente, é composta de textos, fotos, links ou vídeos.

Conforme (ROHR; SQUIRES; PETERS, 2022), os *tweets* possuem um limite máximo de 280 caracteres (140 anteriormente) e são visíveis pelo próprio autor, seguidores ou até mesmo por pessoas interessadas no determinado assunto. Atualmente, a plataforma possui mais de 300 milhões de usuários que se encontram praticamente no mundo inteiro. Além dos *tweets*, os usuários podem se comunicar mutuamente através de *retweets*, mensagens privadas, *tweets* favoritos, menções e/ou hashtags.

Todo usuário da plataforma pode ter *seguidores* e *seguindo*. Chama-se de *seguidores*, usuários que seguem o usuário em questão ou seja, usuários que se inscrevem no perfil de outros usuários para que seja possível visualizar novos conteúdos publicados pelos usuários. Já, *seguindo* são os usuários que o usuário do perfil está seguindo para visualizar ou interagir em conteúdos onde ele não é o usuário principal (SILVA et al., 2015).

A Figura 2 mostra o exemplo de um perfil no microblog com o destaque dos itens mais relevantes. Na parte a), é mostrado o nome oficial inserido pelo usuário. No item b), é exibida a quantidade total de *tweets* publicados pelo usuário na plataforma. O item c) apresenta a foto do perfil. O item d) mostra o nome do usuário formado por um nome escolhido pelo usuário seguido pelo caractere @. Esse nome de usuário deve ser único na plataforma. No item e), são indicados os dados da biografia do usuário. Por fim, nos itens f) e g) respectivamente, é denotada a quantidade de *seguindo* e *seguidores*.

Na Figura 3, é mostrado o exemplo de um *tweet* publicado sobre o Haiti por um ex-comandante das Nações Unidas para a estabilização do Haiti. Na imagem em questão, são destacadas as principais partes de um *tweet*: no item a), da esquerda para a direita, são mostrados, sequencialmente, o nome oficial do perfil, o nome do usuário e a data de publicação do *tweet*; no item b), é exibido o conteúdo oficial do *tweet* com o tema principal em negrito e por fim, no item c), são mostrados os dados de interações de outros usuários ao *tweet* publicado que são,

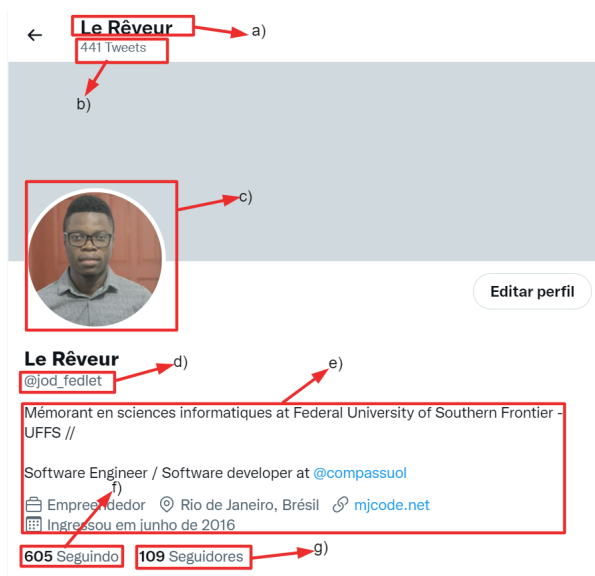


Figura 2 – Exemplo de perfil no Twitter.

respectivamente, quantidade de comentários (28), quantidade de vez que foi retweetado (80), número de curtidas (523) e número de compartilhamento (0) do *tweet*.



Figura 3 – Exemplo de *tweet* publicado sobre o Haiti.

Este capítulo apresentou o domínio da realização da análise exploratória bem como a fonte dos documentos da análise. Tendo o Haiti como foco dos temas dos tópicos a serem analisados, foi fundamental apresentá-lo junto com alguns fatos importantes ocorridos no país ao longo dos tempos. Sendo assim, foi apresentado um breve histórico sobre o país durante o período pré-colonial onde a parte ocidental da ilha *Hispaniola* foi habitada por populações indígenas; o período colonial marcado pela tomada da parte oeste pela França fundamentado no Tratado de Ryswick e pela revolta dos escravos negros, em conjunto aos mulatos, liderada especialmente por Toussaint Louverture e conduziu a independência do Haiti em 1804. Além desses períodos, também, apresentaram-se conceitos sobre crises, instabilidades sociopolítico-econômicas

e emigração. Assim como o Haiti, alguns conceitos sobre o Twitter foram apresentados, sendo que é a mídia social onde os dados da análise serão extraídos. Com isso pretende-se identificar como a comunidade lusófona discute assuntos sobre o Haiti no Twitter.

### 3 MODELAGEM DE TÓPICOS

A abordagem Latent Dirichlet Allocation (LDA), proposta por (DAVID M. BLEI ANDREW Y. NG, 2003), é um dos algoritmos de modelagem probabilística mais utilizados para extrair tópicos em coleções de documentos (CHEHAL; GUPTA; GULATI, 2021). Caracteriza-se por atribuir inicialmente probabilidades às palavras do dicionário encontradas na coleção. A distribuição é feita utilizando a família de distribuição discreta multivariada Dirichlet. Dessa forma, os tópicos são derivados de distribuições probabilísticas de palavras de uma coleção de documentos de entrada. Um conjunto de palavras que, pela relação de ordem, frequência e semântica, representam determinados assuntos (temas). Assim, através dessas relações, é possível definir um tema como um tópico, uma distribuição probabilística de palavras com frequência e semântica que fazem sentido dentro do contexto do tópico.

A Tabela 1 apresenta um exemplo com três tópicos extraídos de sites de notícias e suas cinco palavras principais ao lado das respectivas probabilidades de ocorrência no tópico (coluna  $P(w)$ ). Perceba que, como não há rótulos, o especialista em domínio deve definir a semântica de cada tópico. Por exemplo, o terceiro tópico deve se referir ao início da pandemia na Europa, sendo a palavra *italy* a mais provável de ocorrer (isto é, 2,8%).

A modelagem de tópicos é baseada na ideia de que os documentos são misturas de tópicos, em outras palavras, documentos exibem múltiplos tópicos (STEYVERS; GRIFFITHS, 2007). Assim, os documentos podem ser gerados a partir de diferentes distribuições sobre tópicos. Um documento pode ser definido como uma sequência de palavras  $\mathbf{w} = w_1, w_2, \dots, w_n$ , onde  $n$  é o número de palavras em  $\mathbf{w}$ . Da mesma forma, um corpus (ou coleção) é um conjunto de  $m$  documentos  $D = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$ . Além disso, um documento pode ser qualquer conteúdo baseado em texto, por exemplo, um artigo ou comentário em uma rede social.

Em modelagem de tópicos, a maioria das abordagens considera um documento como um saco de palavras, isto é, a ordem das palavras no documento não importa. Um pré-processamento deve ser realizado na coleção para prepará-la para a extração dos tópicos. A fase de pré-processamento pode ser composta das seguintes etapas (STEYVERS; GRIFFITHS, 2007):

- (i) remoção de *stop words*, ou seja, a remoção de palavras espúrias da coleção;
- (ii) *tokenização*, transformando a coleção em uma lista de palavras;

Tabela 1 – Três tópicos com as suas top-5 palavras e probabilidades.

word	P(w)	word	P(w)	word	P(w)
court	0.025	ventilator	0.028	italy	0.028
police	0.018	mask	0.028	infection	0.015
law	0.015	supply	0.023	lockdown	0.013
prison	0.010	care	0.021	europe	0.012
justice	0.009	doctor	0.019	authority	0.012

- (iii) *stemming*, reduzindo as palavras à sua forma raiz; e
- (iv) lematização, em outras palavras, agrupando as formas inflexíveis de uma palavra.

A Figura 4 representa o modelo LDA (DAVID M. BLEI ANDREW Y. NG, 2003) graficamente. Os retângulos representam iterações: o externo representa os documentos, e o interno representa a escolha repetida de tópicos e palavras dentro de um documento. Além disso, assumindo a LDA como um processo generativo, a Figura 4 pode ser explicado da seguinte maneira:

1. Para cada documento  $w$  da coleção  $\mathcal{D}$ :
  - a) Escolhe-se  $N \sim \text{Poisson}(\xi)$
  - b) Escolhe-se  $\Theta \sim \text{Dir}(\alpha)$
  - c) Para cada  $N$  das palavras  $w_n$ :
    - i. Escolhe-se um tópico  $z_n \sim \text{Multinomial}(\Theta)$
    - ii. Escolhe-se uma palavra  $w_n$  de  $p(w_n | z_n, \beta)$ , probabilidade multinomial condicionada sobre o tópico  $z_n$ .

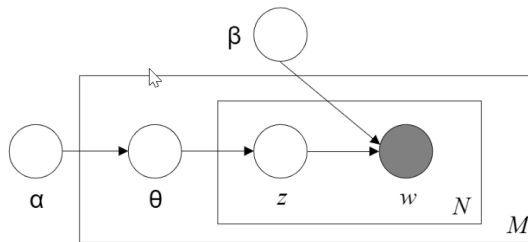


Figura 4 – Representação gráfica do modelo LDA.

O hiperparâmetro  $\beta$  é a contagem de observação prévia sobre o número de vezes que as palavras aparecem em um tópico antes que qualquer palavra do corpus seja observada; um maior valor de  $\beta$  significa que mais palavras estão associadas a um determinado tópico. O hiperparâmetro  $\alpha$  desempenha o mesmo papel, mas em relação aos documentos. Os documentos, como citado anteriormente, podem exibir múltiplos tópicos, assim um mesmo documento pode, por exemplo, tratar sobre política, economia e corrupção. Entretanto, cada tópico associado a documentos tem uma probabilidade diferente. A soma da probabilidade de todos os tópicos associados a um determinado documento é igual a um.

Outra questão quando se trata de modelagem de tópicos é encontrar o número certo de tópicos para uma determinada coleção. Como qualquer método não supervisionado, tem-se que confiar em uma métrica para verificar a melhor combinação dos hiperparâmetros e o número de tópicos (DUARTE; STÅHL, 2018). Tal avaliação poderia ser feita por humanos; entretanto, é uma tarefa custosa e demorada. Os autores (RÖDER; BOTH; HINNEBURG, 2015) apresentam um estudo comparando várias métricas de coerência para modelos temáticos. Este estudo teve

como objetivo descobrir qual métrica é a mais próxima da avaliação humana dos tópicos. A métrica mais correlacionada com a percepção humana foi  $c_v$ .

Baseado no trabalho de (RÖDER; BOTH; HINNEBURG, 2015), a análise exploratória aqui proposta usa  $c_v$  para encontrar o melhor número de tópicos ( $K$ ). Para a melhor combinação de  $\alpha$  e  $\beta$  foi definida com o valor *auto* do modelo, isto é, o modelo aprende automaticamente os melhores valores para os hiperparâmetros à medida que é treinado no corpus.



#### 4 TRABALHOS RELACIONADOS

Neste capítulo, serão apresentados os quatro trabalhos relacionados ao tema que foram selecionados a fim de direcionar o estudo. Esses trabalhos são de (PEREIRA, 2019), (HIDAYA-TULLAH et al., 2018), (ASGHARI; SIERRA-SOSA; ELMAGHRABY, 2018) e (OH; KWON; RAO, 2010).

No trabalho de (PEREIRA, 2019), é realizada uma análise exploratória de *tweets* extraídos na base de dados do Twitter, utilizando o modelo *BTM (Biterm Topic Model)*, tendo como objetivo geral descobrir os tópicos discutidos durante o período dos Jogos Olímpicos de 2016. Alguns dos objetivos específicos do trabalho eram:

- definir o método a ser utilizado;
- definir o período de datas para a obtenção dos *tweets*;
- realizar o pré-processamento dos dados;
- verificar e analisar os principais assuntos referentes aos Jogos Olímpicos Rio 2016.

Para extrair os *tweets* na base do Twitter, foram utilizadas as seguintes strings de busca: "rio2016", "olimpiadas", "olimpiada", "ceremoniadeabertura", "ceremoniadeencerramento", "jogosolimpicos", "olympics", "olympicgames", "openingceremony", "closingceremony". Foi definido também, o período de 02 de agosto de 2016 a 24 de agosto de 2016 para realizar as devidas buscas. Como resultado, foi obtido um total de 2.806.785 *tweets* divididos em 230 arquivos, ou seja, um arquivo para cada palavra da *string* de busca (10) multiplicado pela quantidade de dias (23). Em seguida, os arquivos foram divididos em grupos de 9, um para cada intervalo definido.

A Figura 5 apresenta o conjunto de *tweets* obtidos em que os *tweets* extraídos na base do Twitter são divididos em 9 arquivos e agrupados em fatias de dias.

Arquivo	Dias	Tamanho	Quantidade de Tweets
1	02.08.2016 a 04.08.2016	27,2MB	128.527
2	05.08.2016	65,9MB	334.125
3	06.08.2016 a 08.08.2016	127MB	624.444
4	09.08.2016 a 11.08.2016	51MB	239.912
5	12.08.2016 a 14.08.2016	52,1MB	245.149
6	15.08.2016 a 17.08.2016	64,7MB	311.581
7	18.08.2016 a 20.08.2016	59,2MB	279.432
8	21.08.2016	53,9MB	260.063
9	22.08.2016 a 24.08.2016	74MB	350.763

Figura 5 – Exemplo de *tweets* extraídos na base de dados do Twitter.

Fonte: (PEREIRA, 2019)

Tendo os *tweets* extraídos, divididos em arquivos e agrupados por intervalo de dias, foi realizado o pré-processamento dos dados, processo em que é realizada a limpeza dos dados. Para isso, foi necessário remover e/ou tratar alguns dados julgados desnecessários na continuação da análise exploratória. Os tratamentos realizados foram os seguintes:

- Conversão dos *tweets* em letras minúsculas;
- Remoção de *stop words*, pontuação e acentuação, e das URLs;
- Remoção de *tweets* contendo menos de 3 palavras, exceto quando estivesse representando país, por exemplo: br, fr, us e de *tweets* duplicados.
- Remoção de menções (palavras seguidas do caractere '@') e de *hashtags* (palavras seguidas do caractere '#');
- Remoção de dígitos com exceção de 2016 quando seguido da palavra rio.

Após o pré-processamento dos *tweets*, reduziu-se o conjunto de 2.806.785 *tweets* para 1.548.759, com um total de 8.611.136 palavras e com uma média de 5,56 palavras por *tweet*. Posteriormente ao pré-processamento dos dados, foram aplicado o algoritmo de modelagem de tópicos *BTM* para identificar os tópicos na coleção de documentos e realizado o pós-processamento dos dados onde foi definida a quantidade de tópicos que será utilizada no experimento, sendo K, número de tópicos, igual a 5, 10, 15, 20 e 30.

Na Figura 6 é mostrado o exemplo de um *tweet* obtido após a extração (lado esquerdo) contendo URL, caracteres especiais como '#' e '@' e o mesmo após a aplicação da etapa de pré-processamento dos dados (lado direito).

Original	Após limpeza
Dia de futebol feminino nessas Olimpíadas. #Rio2016 @Arena Corinthians - Itaquerao <a href="https://www.instagram.com/p/BIyVNoOBHvD/">https://www.instagram.com/p/BIyVNoOBHvD/</a>	futebol feminino olimpiadas corinthians itaquerao

Figura 6 – Exemplo de *tweets* antes e depois do pré-processamento dos dados.

Fonte: (PEREIRA, 2019)

Como resultados do trabalho, cada arquivo de *tweets* da Figura 5 é representado por uma tabela contendo os tópicos e as suas probabilidades de ocorrência. Em cada tabela, os tópicos estão em ordem crescente, ou seja, o tópico mais provável aparece na primeira posição, seguidos da sua probabilidade de ocorrência e o conjunto de palavras relacionado ao determinado tópico. Depois disso, foi realizada a rotulação dos tópicos manualmente baseada em fatos empíricos.

A Figura 7 é uma representação do arquivo 1 da Figura 5 dos tópicos extraídos do dia 05 de agosto de 2016. É possível notar que muitas das palavras se referem a momentos

ocorridos no evento, como exemplo, palavras como 'anita', 'caetano', 'gil', 'cantando', 'cantar', 'mc', 'fernanda', 'karol', 'montenegro', 'gilberto', mencionam as seguintes apresentações: Anita, Caetano Veloso e Gilberto Gil cantando músicas de Ary Barroso e João Gilberto, Karol Conka e Mc Soffia cantando rap e Fernanda Montenegro declamando trechos de poema "A Flor e a Náusea", de Carlos Drummons de Andrade. Por outro lado, as palavras como 'pira', 'olimpica', 'acender', 'vanderlei', 'tocha', 'guga', 'cordeiro', 'lima' e 'chama', falam do momento em que a chama olímpica aparece no Maracanã nas mãos do ex-tenista Guga, e em seguida, o ex-maratonista Vanderlei Cordeiro de Lima a recebe, acendendo a pira e oficializando o início dos Jogos Olímpicos Rio 2016 (PEREIRA, 2019).

Rótulo	Palavras
Cerimônia de Abertura	país, lindo, copa, mundo, brasileiro, povo, dinheiro, bonito, mal, festa, deus, gisele, demais, cerimonia, orgulho, maravilhosa, amo, parabens, país, melhor, hoje, momento, atleta, anitta, caetano, gil, cantando, cantar, mc, fernanda, karol, montenegro, gilberto, jogo, medalha, ouro, futebol, esporte, selecao, olimpico, delegacao, bandeira, alemanha, mulher, vem, porta, grecia, aquecimento, indios, portugueses, global, parte, mostrar, historia, hino, musica, paulinho, viola, nacional, zeca, flamengo, pira, olimpica, acender, vanderlei, tocha, guga, cordeiro, pele, lima, chama, maracana, janeiro, cidade, estado, países, samba, aula, carnaval, geografia, escola, falar, regina, ingles, case, falando, portugues, tremendo, santos, fala, homem.
Mídia	galvao, boca, cala, gloria, maria, globo, ouvir, bueno, falando, falar.
Político	temer, vaia, dilma, presidente, vaiado, golpista, michel, lula, medo, povo.
Transmissão das Olimpíadas	pokemon, vendo, hoje, cerimonia, mundo, casa, assistir, assistindo, estar, tv.

Figura 7 – Tópicos obtidos a partir dos *tweets* extraídos do dia 05 de Agosto de 2016

Fonte: (PEREIRA, 2019)

Em (HIDAYATULLAH et al., 2018), foi utilizada a modelagem de tópicos para determinar tópicos oriundos de *tweets* das competições de futebol escritos em Bahasa (idioma oficial da Indonésia). Foram escolhidas 5 ligas: La Liga nacional da Indonésia, a Premier League da Inglaterra, a Bundesliga da Alemanha, a La Liga da Espanha e a Série A da Itália. Os dados do estudo foram coletados em várias contas oficiais do Twitter indonésio que, em todo o tempo, atualizam sobre o futebol. Para definir o tipo dos tópicos, foi aplicado o Latent Dirichlet Allocation (LDA) como método de modelagem de tópicos.

Para obter os *tweets* na base de dados do Twitter sobre as competições supracitadas, usaram-se cinco contas oficiais que são @VIVAbola, @panditfootball, @detiksport, @Bolanet e @GOAL\_ID e definiu-se o intervalo de tempo entre o dia 01 de janeiro de 2017 e 24 de dezembro de 2017. Com isso, foi obtido um total de 120.639 *tweets*. Após a obtenção dos *tweets*, foi realizado o pré-processamento com as seguintes etapas:

- Remoção de tags HTML e caracteres Unicode;
- Remoção de símbolos e emoticons;
- Remoção de caracteres especiais e URLs; e



Ao aplicar o LDA, constatou-se que cinco foi o melhor número de tópicos a serem extraídos.

Como resultados, cada tópico obtido possui porcentagens de interesse maior em determinados estados e menor em outros. Por exemplo, para o tópico *Care*, Califórnia e Nova York possuem uma porcentagem de 80% de interesse, ou seja são os estados mais interessados e Virgínia com 41% é o menos interessado.

O trabalho de (OH; KWON; RAO, 2010) propõe uma exploração da dinâmica de trabalho de fábrica de boatos (*fake news*) analisando dados do Twitter tendo como foco os *tweets* postados durante o terremoto no Haiti em 2010. Foram realizados processo de identificação dos tipos de conteúdo das postagens dos *tweets* por meio de codificação e processo de análises quantitativas e qualitativas. Como esquema de codificação, foi usado o Rumor Interaction Analysis Systems (RIAS), ANOVA de Friedman como método estatístico não paramétrico para a análise quantitativa na qual é estudada a mudança de padrões de modo de comunicação e o Automap, software de visualização usado na análise de rede semântica que é uma análise qualitativa.

Para coletar os dados relacionados ao terremoto de 2010 no Haiti no microblog Twitter, usou-se *#haitiearthquake* como palavra-chave e definiu o intervalo de dez (10) dias, partindo do primeiro dia do terremoto, ou seja do dia 12 de janeiro de 2010 até o dia 21 de janeiro de 2010. Após essas definições, obteve-se um total de mais de 3.000 *tweets* por dia contendo *tweets* não relacionados ao assunto principal e *tweets* postados em idiomas francês e crioulo que não são relevantes para a pesquisa. Portanto, os mesmos foram excluídos e ficou com um total de 960 *tweets* relacionados ao tema em língua inglesa. Posteriormente, foram realizadas limpeza de dados e análises quantitativas, definindo o modelo de quatro estágios como o mais adequado.

Consequentemente, ao executar a rede semântica com os quatro conjuntos de dados igualmente divididos e identificados precedentemente para explorar padrões de mudança no modo de comunicação, partindo de amostras de cada estágio, as palavras como *blog*, *CNN*, *picture*, *list*, *info*, *report* e *update* são relacionadas a declarações de autenticação e são um exemplo dos resultados obtidos.

Este trabalho, como todos os citados neste capítulo, utilizou *tweets* para gerar a coleção de documentos. Conforme já apresentado, os *tweets* são relacionados ao país Haiti (como o trabalho de (OH; KWON; RAO, 2010)) porém utilizando a abordagem LDA para a descoberta dos tópicos (como em (HIDAYATULLAH et al., 2018) e (ASGHARI; SIERRA-SOSA; ELMAGHRABY, 2018)).

## 5 METODOLOGIA

Neste capítulo, serão apresentados os materiais (corpus e ferramentas) e método (LDA) usados na identificação dos assuntos mais discutidos sobre o Haiti no Twitter no período de janeiro de 2018 a maio de 2022. A metodologia foi realizada em quatro etapas que são: a coleta dos dados, o pré-processamento dos dados, a aplicação do modelo LDA e a rotulação manual dos tópicos.

### 5.1 FERRAMENTAS

A parte da extração dos *tweets* foi realizada em ambiente local e a parte do pré-processamento dos dados e aplicação do modelo LDA foi realizada no *Google Collabotary (Colab)*<sup>1</sup>, notebook que permite a execução de código Python<sup>2</sup> em nuvem e que possui máquinas robustas. Essa separação se deveu a algumas limitações encontradas em questão de versão de Python usada no Colab (3.7). Sendo que a ferramenta de extração dos *tweets* requer versões acima de 3.7 para poder usufruir de todas as funcionalidades disponibilizadas. Para o ambiente local, utilizou-se um computador de *Sistema Operacional Windows 10 Pro-64 bits, de processador 11th Gen Intel (R) Core (TM) i7-1185G7 @ 3.00GHz 1.80GHz e de memória RAM 16GB*. Já para o Colab, usou-se uma GPU NVIDIA Tesla T4 com 32 GB de memória. Outras ferramentas que também foram utilizadas estão listadas a seguir:

- Snsrape<sup>3</sup>: biblioteca que não necessita nenhum tipo de autenticação, utilizada para coletar os dados no Twitter;
- Pandas<sup>4</sup>: biblioteca utilizada para ler e manipular os dados de entrada;
- NLTK<sup>5</sup>: biblioteca utilizada para remover as *stopwords* em língua portuguesa;
- Gensim<sup>6</sup>: biblioteca da linguagem Python utilizada para treinar o modelo LDA;
- Matplotlib<sup>7</sup>: biblioteca da linguagem Python utilizada para visualizar os dados obtidos.

### 5.2 COLETA DOS DADOS

Neste trabalho, foi usada a rede social Twitter para coletar as informações por ser aberta e muito acessível. Ela possui uma grande quantidade de internautas que compartilham

<sup>1</sup> <https://colab.research.google.com/>

<sup>2</sup> <https://www.python.org/>

<sup>3</sup> <https://github.com/JustAnotherArchivist/snsrape>

<sup>4</sup> <https://pandas.pydata.org/>

<sup>5</sup> <https://www.nltk.org/>

<sup>6</sup> <https://radimrehurek.com/gensim/>

<sup>7</sup> <https://matplotlib.org/>

informações constantemente. Assim, o módulo do Twitter da biblioteca *Snsrape* foi utilizado para coletar os dados nos períodos de datas definidas entre o dia 01 de janeiro de 2018 ao dia 31 de maio de 2022. Outros parâmetros foram usados como palavras-chave: *haiti*, *haitiano* e *haitiana*, filtrando apenas os *tweets* escritos em língua portuguesa, com a opção *lang:pt* da ferramenta *Snsrape*. Assim, foi obtido um total de 185.348 *tweets* contendo 4.542.561 palavras, sendo uma média de 24,51 de palavras por *tweet*.

### 5.3 PRÉ-PROCESSAMENTO DOS DADOS

Como nos trabalhos relacionados, foi utilizado o pré-processamento dos dados neste trabalho. Processo em que são tratados os dados que não apresentaram uma boa semântica para o modelo de tópicos. Os tratamentos realizados eram:

- Conversão dos textos em letras minúsculas;
- Remoção de dígitos, URLs, *stop words*, pontuação e acentuação;
- Remoção de palavras que:
  1. iniciam com caracteres @ (menções) ou # (hashtags).
  2. possuem menos de três caracteres e são diferentes de *ht* (sigla do país Haiti).
  3. possuem mais de três ocorrências seguidas (Por exemplo, kkkkkk).
- Cálculo de bigramas, técnica que faz com que palavras que aparecem juntas com uma certa frequência fiquem juntas. Por exemplo, palavras como Rio de Janeiro são transformadas para *rio\_janeiro*. No cenário deste trabalho, foi considerado três como a frequência de aparição;
- Eliminação de *tweets* duplicados, mantendo somente o primeiro;
- Eliminação de *tweets* com menos de três palavras;
- Aplicação de lematização, utilizada também no trabalho de (ASGHARI; SIERRA-SOSA; ELMAGHRABY, 2018) apresentado no capítulo 4.

A Tabela 2 mostra o exemplo de um *tweet* antes e depois do processo de pré-processamento. Perceba que após o pré-processamento, as palavras iniciadas com cerquilha (#), links ou outros dados indesejados foram descartados conforme definidos anteriormente.

Após o pré-processamento, o número de *tweets* foi reduzido para 150.174 e 2.127.695 de palavras no total. A média de número de palavras agora é de 14,17. A Tabela 3 mostra a quantidade de *tweets*, de palavras e a média das palavras antes e depois do pré-processamento dos dados, agrupadas por ano. Perceba que 2022 é o ano com a menor ocorrência, pois compreende apenas cinco meses.

<b>Antes do pré-processamento</b>
FUGA DO BOZO: 'Presidente, o senhor é racista?', pergunta repórter a Trump sobre polêmica envolvendo Haiti e países africanos ele saiu sem responder. <a href="https://t.co/AUprDvBNT0">https://t.co/AUprDvBNT0</a>
<b>Depois do pré-processamento</b>
fuga bozo presidente senhor racista perguntar reporter trump sobre polemica envolver haiti pais africano sair responder

Tabela 2 – Exemplo de *tweet* antes e depois do pré-processamento dos dados.

<b>Antes do pré-processamento</b>			
<b>Ano</b>	<b>Quantidade de Tweets</b>	<b>Quantidade de Palavras</b>	<b>Média de Palavras</b>
<b>2018</b>	23.691	531.047	22.42
<b>2019</b>	29.739	742.523	24.97
<b>2020</b>	56.375	1.317.064	23.36
<b>2021</b>	62.625	1.621.237	25.89
<b>2022</b>	12.918	330.690	25.60
<b>Depois do pré-processamento</b>			
<b>Ano</b>	<b>Quantidade de Tweets</b>	<b>Quantidade de Palavras</b>	<b>Média de Palavras</b>
<b>2018</b>	18.191	246.135	13.53
<b>2019</b>	24.437	363.393	14.87
<b>2020</b>	47.500	647.063	13.62
<b>2021</b>	48.707	701.327	14.40
<b>2022</b>	11.339	169.777	14.97

Tabela 3 – Repartição anual dos *tweets* antes e depois do pré-processamento.

## 5.4 APLICAÇÃO DO MODELO LDA

A aplicação do modelo LDA foi dividida nas seguintes etapas: treinamento do modelo e identificação do melhor número de tópicos, extração dos tópicos e cálculo dos tópicos mais frequentes de cada mês.

### 5.4.1 Treinamento do modelo e identificação do melhor número de tópicos

Nesta etapa, realizou-se o treinamento do modelo LDA usando a biblioteca *Gensim*. Os dados coletados e pré-processados foram repartidos em 53 arquivos, ou seja, um arquivo para cada mês de janeiro de 2018 a maio de 2022. Isso ajudou a obter tópicos mais precisos e concisos, facilitando a contextualização dos resultados. Para cada mês, definiu-se aleatoriamente o número de tópicos 3, 5, 7, 10, 15 e 25, denotado por  $K$ . Além disso, os dados pré-processados de cada mês foram transformados em uma forma vetorizada, *Bag of words* (sacola de palavra), calculando a frequência de cada palavra na coleção. Por exemplo, para um documento  $\mathbf{w}=\{w_1, \dots, w_n\}$ , a sua transformação vetorial ficou o formato  $[(id_1, m_1), \dots (id_n, m_n)]$ , sendo  $id_i$  o identificador de  $w_i$  e  $m_i$  a quantidade de vez que  $id_i$  aparece no documento. Para que essa transformação fosse possível, primeiramente criou-se um dicionário que representa as coleções de documentos. Ao criar o dicionário, os parâmetros *no\_below* com valor de 10% e *no\_above* com valor de 80% foram definidos. Isso é, palavras que aparecem em menos de 10% ou em mais de 80% dos



documentos foram descartadas.

Tendo os números de tópicos definidos e os documentos transformados, realizou-se a configuração do modelo com os seguintes parâmetros:

- Proporção dos tópicos no documento ( $\alpha$ ) e proporção das palavras nos tópicos ( $\beta$ ). Ambas foram configuradas com o valor *auto*, isto quer dizer que o modelo aprende automaticamente os melhores valores para os hiperparâmetros conforme ele é treinado no corpus.
- Controlador da frequência com que o modelo foi treinado no corpus (conhecido como *passes*) foi definido para 500;
- Número de iterações, frequência com que um *loop* específico é repetido em cada documento, foi definido para 1000.

Posteriormente às configurações, aplicou-se o modelo LDA com todos os números de tópicos definidos anteriormente. Como resultados, retornou os tópicos de cada  $K$  de cada mês. Logo, usou-se a métrica de coerência  $c_v$  da *Gensim* para identificar o melhor número de tópicos de cada mês. O melhor número de tópicos de cada mês é o  $K$  que apresentou o maior valor de  $c_v$ . Portanto, para o melhor número de tópicos de toda a experimentação, considerou-se aquele que mais se repete, sendo 5 esse número.

Logo, percebeu-se que as palavras como *nao, para, pra, aqui, ir, ter, esse, todo, tudo, este, ser, querer, sao, pai, pro, ver, voce, sobre, outro, sim, agora, de, aler, de\_esse, ate, em\_este, deu, em\_esse, vcs, em\_o, tao, cara, estao, vao, nome, tamper, segundo, colomber, serar, carar, onde, assim, dia, ainda, vez, dois, apo, hoje, pois, algum, etc, exemplo, algum, um, mesmo, enquanto, tipo, tanto, tanta, muito, sempre, mano, bbb, tar, kkk, algo, pouco, tbm, quase, porque, nunca, nada, desde, cade, tal, de\_isso, tr, fazer, dizer, falar, bem, bom, dar, vir, achar, saber, ninguem, poder, dever, nenhum, novo, quanto, talvez* não ajudavam no entendimento dos tópicos. Por isto, foram descartadas e todo o processo de treinamento foi repetido.

#### 5.4.2 Extração dos tópicos e identificação dos mais frequentes em cada mês

Uma vez ter o modelo treinado e o melhor número de tópicos definido, aplicou-se o modelo LDA novamente para retornar os tópicos da experimentação com  $K = 5$ . Considerando também as demais configurações do modelo que foram definidas na etapa de treinamento como número de passes, de iterações e os valores dos hiperparâmetros  $\alpha$  e  $\beta$ . Por conseguinte, os tópicos foram extraídos para cada mês, de janeiro de 2018 a maio de 2022.

Logo após a extração dos tópicos, realizou-se o cálculo dos tópicos mais frequentes nos documentos de cada mês. Esse cálculo ajuda a descobrir a relação dos documentos com cada tópico, ou seja, a importância de cada tópico para os documentos. Para tal, o método *get\_document\_topics* implementado no modelo LDA da biblioteca *Gensim* foi utilizado. Como parâmetro, utilizou-se o *minimum\_probability* com o valor de 0.33. Isto é, para cada documento,

foi retornada uma distribuição de tópicos, considerando somente os tópicos que representavam pelo menos 33% como probabilidade no documento. Sendo 5 o número de tópicos, considerou-se, no máximo, três tópicos para cada documento, ou seja, a probabilidade de 33%.

## 5.5 ROTULAÇÃO MANUAL DOS TÓPICOS

Esta etapa consiste em atribuir um tema (assunto) aos tópicos obtidos baseando na interpretação das palavras desses tópicos. No entanto, em modelagem de tópicos, não existe nenhuma forma apropriada ou métodos científicos definidos para realizar tal processo. Geralmente, pode-se basear em domínio da *coleção* (dados do estudo) apresentado pelo autor.

Desse modo, neste trabalho, a rotulação foi baseada nas seguintes etapas: (i) conhecimentos empíricos da *coleção*, (ii) busca por notícias no *Google* <sup>8</sup> usando as dez palavras mais importantes em cada tópico, filtrando por faixa de datas dos tópicos, ou (iii) a leitura dos *tweets* originais baseando nas palavras mais importantes. A coleção original foi dividida em 53 subcoleções (53 meses), resultando, assim, em 265 tópicos ( $53 \times 5$ ). Como a rotulação manual é dispendiosa, foi rotulado apenas o tópico que está associado a mais documentos usando uma das etapas supracitadas.

A Figura 9 mostra um exemplo do processo de rotulação de tópicos, baseado em notícias do *Google*. Observa-se que foram usadas as palavras mais frequentes do tópico que são: *haiti, trump, brasil, eua, africa, pessoa, brasileiro, chamar, onu, terremoto* e filtragem por faixa de datas entre o dia 01 de janeiro de 2018 a 31 de janeiro de 2018, período no qual o tópico foi extraído. Assim, retornaram-se notícias relacionadas ao tópico com a sua determinada faixa de datas. Ao ter as notícias, realizou-se leitura para entender o contexto do tópico nas notícias. Por conseguinte, a rotulação mais apropriada a esse tópico é *Imigração e racismo*.



Figura 9 – Exemplo de rotulação usando notícias do *Google*.

<sup>8</sup> <https://www.google.com/>

## 6 RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados obtidos com a metodologia empregada e as discussões desses resultados. Os resultados são discutidos por mês de cada ano dos *tweets* extraídos, ou seja, de janeiro de 2018 a maio de 2022.

### 6.1 2018

A Tabela 4 apresenta os top-1 tópicos de cada mês do ano de 2018 em que cada tópico está representado pelas suas 10 respectivas palavras mais relevantes e a rotulação manual. A ordem de aparição das palavras depende da importância de cada uma para o tópico, ou seja, a palavra com maior probabilidade aparece primeiro e aquela com menor probabilidade aparece por último. O ano de 2018 iniciou com temas sobre *Imigração e racismo*, finalizando com temas relacionados com *Imigração e educação*. Sendo que o tema mais frequente durante o ano foi *Intervenção militar*. Observando a Tabela 4, percebe-se que esse tópico aparece como mais frequente nos meses de fevereiro, março, abril, agosto e setembro. O Apêndice A apresenta todos os cinco tópicos de cada mês.

Mês	Palavras	Assunto
jan.	haiti, trump, brasil, eua, africa, pessoa, brasileiro, chamar, onu, terremoto	imigração e racismo
fev.	exercito, militar, general, brasileiro, povo, rio_janeiro, trabalho, usar, coisa, sexual	intervenção militar
mar.	haiti, brasil, rio, gente, exercito, brasileiro, militar, sirio, ajudar, venezuela	intervenção militar
abr.	haiti, brasil, exercito, usar, militar, vida, missao_paz, terremoto, mandar, missao	intervenção militar
maio	contra, messi, selecao, jogo, grande, cuba, argentino, amistoso, fazer_gol, hat_trick	jogo de futebol
jun.	haiti, brasil, brasileiro, selecao, povo, ficar, ajudar, futebol, pessoa, venezuela	jogo de futebol
jul.	haiti, gente, brasileiro, jamaica, ficar, selecao, chorar, pessoa, mundo, pais	jogo de futebol
ago.	haiti, brasil, militar, brasileiro, ficar, chegar, bolsonaro, missao, mundo, grande	intervenção militar
set.	haiti, brasileiro, passar, pessoa, missao_paz, contra, governo, mundo, missao, militar	intervenção militar
out.	haiti, povo, cuba, brasileiro, passar, militar, presidente, paz, conhecer, guerra	comparação econômica
nov.	haiti, brasil, grande, ajudar, mundo, contra, eua, usar, coisa, deixar	imigração
dez.	haiti, brasil, militar, brasileiro, imigrante, bolsonaro, haitiana_cega, governo, povo, aprovar_oab	imigração e educação

Tabela 4 – Tópico predominante de cada mês de 2018.

Os meses de janeiro, outubro, novembro e dezembro foram marcados, respectivamente, por *Imigração e racismo*, *comparação econômica*, *Imigração e Imigração e educação*. Em janeiro, o assunto relata as falas do ex-presidente dos Estados Unidos, Donald Trump, em que ele teria se referido, em reunião na *Casa Blanca* com legisladores, ao Haiti, El-Salvador, países da América Central e aos países africanos como "*países de merda*" e sugeriu que os Estados Unidos recebessem imigrantes da Noruega ao invés daqueles vindos dos países suprarreferidos. Por exemplo, "*Trump sobre imigrantes de El Salvador e Haiti: "Por que recebemos pessoas de países de merda?"*" <https://t.co/61a8PkCgbH> <https://t.co/SpG7bT8Jl1>" representa o extrato de um *tweet* associado a esse tópico. Em outubro, o povo brasileiro ficou angustiado após as eleições presidenciais que elegeram o Jair Messias Bolsonaro, militar de profissão. A angústia aparece quando o povo disse que não queria que a economia brasileira tornasse aquela do Haiti ou a da Venezuela. Em novembro, a discussão era em torno da imigração onde os brasileiros

achavam normal o fato de eles terem emigrado ilegalmente para os Estados Unidos mas, a vinda dos haitianos para o Brasil era totalmente intolerável mesmo tendo visto. Já o mês de dezembro foi marcado pelas discussões a respeito de uma imigrante haitiana, cega, aprovada na Ordem dos Advogados do Brasil (OAB), que queria se naturalizar e virar juíza.

Nos períodos de fevereiro, março, abril, agosto e setembro, percebe-se a dominação do tema *Intervenção militar*. Isto se refere à atuação do exército brasileiro, durante vários anos, como força de paz das Organizações das Nações Unidas (ONU) no Haiti. Da mesma forma, se refere à intervenção da força armada brasileira que ocorreu no Rio de Janeiro e que teve como objetivo principal a diminuição da taxa de criminalidade. Nessa intervenção, o general que também atuou no Haiti fez comparações entre a atuação do exército no Haiti e no Rio de Janeiro. Identicamente, foram vários outros comentários de cidadãos brasileiros comparando o que aconteceu no Rio de Janeiro com o que já aconteceu no Haiti em questão de guerra civil para entender as diferenças e similaridades.

Maio, junho e julho foram destacados por assunto sobre futebol. Principalmente, pelo jogo da seleção argentina de futebol que venceu por 4 a 0 a seleção haitiana, em amistoso internacional no dia 25 de maio de 2018. Nesse jogo, os artilheiros Lionel Messi e Sérgio Agüero marcaram, respectivamente, 3 (*hat trick*) e 1 gols. Também foi marcado pela derrota da seleção brasileira pela seleção belga, na Copa do Mundo de futebol no dia 06 de julho de 2018, cuja a seleção brasileira possui uma grande torcida no Haiti. Antes do jogo, o povo haitiano disfarçou que tinha esquecido dos problemas sociopolítico-econômicos pelos quais o país passava, mas, a eliminação do Brasil ajudou a deflagrar uma crise que levou à queda do primeiro-ministro haitiano, Jack Guy Lafontant.

A Figura 10 apresenta a porcentagem do tópico mais frequente em relação aos demais de cada mês do ano de 2018. Perceba que em oito de 12 meses, um tópico é dominante chegando a estar associado a mais de 50% dos *tweets*. Por exemplo, os meses de julho e junho foram dominados pelo assunto de *Jogo de futebol* (conforme a Tabela 4) tendo 91.26% e 88.19% dos *tweets* associados a eles, respectivamente.

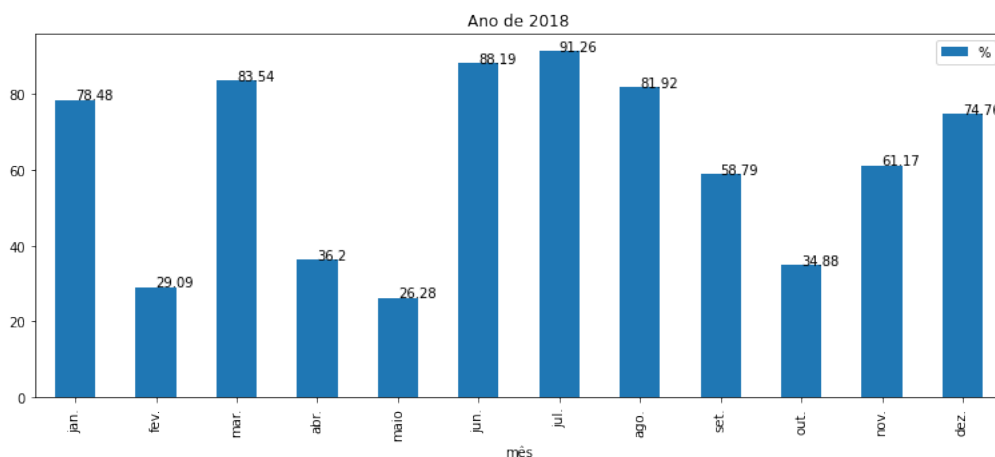


Figura 10 – Frequência dos tópicos predominantes de cada mês de 2018.

## 6.2 2019

O ano de 2019 iniciou com assunto relacionado ao *Terremoto de 12 de jan. de 2010* e terminou com tema referente à *Intervenção militar*. A Tabela 5 mostra as 10 palavras (por ordem decrescente de probabilidade das palavras) de cada tópico obtido durante o ano de 2019 com as respectivas rotulações manuais. Nesse ano, o tópico *Ajuda humanitária discriminativa* aparece como mais frequente nos meses de fevereiro, março, abril e maio (em conformidade à Tabela 5). O Apêndice B apresenta todos os cinco tópicos de cada mês.

Mês	Palavras	Assunto
jan.	haiti, terremoto, ajudar, brasileiro, exercito_brasileiro, militar, brumadinho, eua, atuar, precisar	terremoto de 12 de jan. de 2010
fev.	eua, petroleo, venezuela, humanitario, ajuda_humanitaria, ajudar_humanitaria, mundo, trump, mandar, pais	ajuda humanitária discriminativa
mar.	haiti, brasil, venezuela, eua, ajudar, governo, petroleo, cuba, grande, mundo	ajuda humanitária discriminativa
abr.	haiti, pessoa, gente, ajudar, ficar, mundo, brasileiro, mocambique, venezuela, coisa	ajuda humanitária discriminativa
maio	haiti, brasil, brasileiro, venezuela, ficar, jovem, ajudar, grande, gente, governo	ajuda humanitária discriminativa
jun.	brasil, costa_rica, copa_ouro, grupo, jogo, eua, nicaragua, copa_america, curacao, contra	jogo de futebol
jul.	haiti, brasil, mexico, virar, jogo, eua, copa_ouro, ouro, grande, perder	jogo de futebol
ago.	haiti, brasil, brasileiro, chamar, presidente, mundo, governo, pobre, povo, general	comparação sociopolítico-econômica
set.	haiti, general, brasileiro, pessoa, povo, deixar, contra, mundo, pais, coisa	intervenção militar
out.	haiti, brasil, equador, protesto, presidente, mundo, chile, gente, rua, hong_kong	protesto contra o governo haitiano
nov.	haiti, heleno, contra, eua, negro, cuba, mundo, pensar, pais, pobre	atuação do general Heleno no Haiti
dez.	haiti, brasil, onu, soldado, filho, mulher, missao, ajudar, pessoa, vida	intervenção militar

Tabela 5 – Tópico predominante de cada mês de 2019.

Janeiro, agosto, outubro e novembro foram marcados, respectivamente, por *Terremoto de 12 de jan. de 2010*, *Comparação sociopolítico-econômica*, *Protesto contra o governo haitiano* e *atuação do general Heleno no Haiti*. Em janeiro, as discussões eram mais em torno do terremoto que devastou o Haiti no dia 12 de janeiro de 2010. Nesse tremor, vários cidadãos brasileiros, em ajuda ao país mediante a atuação do exército brasileiro, foram mortos. No mês de agosto, os internautas fizeram comparações do governo do presidente Bolsonaro em relação a um regime ditatorial. Através dessas comparações, percebeu-se a preocupação do povo que não queria que o Brasil virasse o Haiti, colocando-o na pobreza. O mês de outubro referiu-se às manifestações do povo haitiano que ocupou as ruas reclamando da má gestão do governo haitiano. Já o mês de novembro foi notado pelas conversas contra o general Augusto Heleno, ministro da defesa do governo do Bolsonaro e ex-general da intervenção militar do Haiti.

Em fevereiro, março, abril e maio o assunto mais destacado era *Ajuda humanitária discriminativa*. Nos meses de fevereiro e março, por exemplo, a discussão era em torno do governo dos Estados Unidos que mostrou interesse e preocupação pela Venezuela e não pelo Haiti apesar das dificuldades que esse último enfrenta há muito mais tempo. Percebeu-se que o interesse dos Estados Unidos estava meramente no petróleo da Venezuela. Pode-se observar o extrato de um *tweet* referente a esse tema: "*@LeoJaime Quais informações veridicas que temos sobre a Venezuela? Muita gente querendo sair do país, isso não é privilégio só de lá. Quais os reais motivos da crise econômica? Apenas da má gestão? Só sei que sem petróleo o mundo estaria se lixando para a Venezuela. Quem liga pro Haiti?*". Em abril, os comentários falam da preocupação da comunidade internacional pela França, posteriormente ao incêndio da catedral da Nossa Senhora de Paris. O mundo se mobilizou em ajudar a França ao invés do Haiti,

Moçambique e outros países que estavam passando por situações piores. No entanto, o Brasil é um dos país que sempre ajudou incondicionalmente esses países que passam por necessidades.

Os meses de junho e julho foram destacados pelo assunto de *Jogo de futebol*. Especificamente, as discussões referenciam a participação da seleção haitiana de futebol na Copa de Ouro de 2019, organizada pela Confederação de Futebol da América do Norte, Central e Caribe (CONCACAF). No mês de junho, os comentários eram referentes à divisão dos grupos em que o Haiti ficou no Grupo B com Costa Rica, Nicarágua e Bermudas. Em julho, a discussão era sobre a derrota da seleção haitiana de futebol pela seleção mexicana na semifinal da competição.

Como no ano de 2018, o assunto *Intervenção militar* apareceu no ano de 2019 e dominou os meses de setembro e dezembro. Particularmente, esses períodos foram marcados por comentários referentes à atuação do exército brasileiro no Haiti. Em setembro, as falas foram sobre o pedido da ONU para que o general Heleno deixasse o comando da missão por comandar um massacre de pessoas inocentes na comunidade de Cité Soleil. O mês de dezembro relata as lembranças compartilhadas por ex-soldados ou outros profissionais como enfermeiros que participaram na missão.

Dentre os tópicos do ano de 2019, percebe-se pela Figura 11, que os meses de setembro e outubro representados pelos assuntos *Intervenção militar* e *Protesto contra o governo haitiano* (de acordo com a Tabela 5), respectivamente, apresentavam as maiores porcentagens como tópicos mais frequentes. Sejam os valores de 93.06% e 90.46%. Percebe-se, também, que as discussões no Twitter são bem polarizadas para um assunto, pois nove dos 12 meses um assunto (tópico) está associado a mais de 50% dos *tweets*.

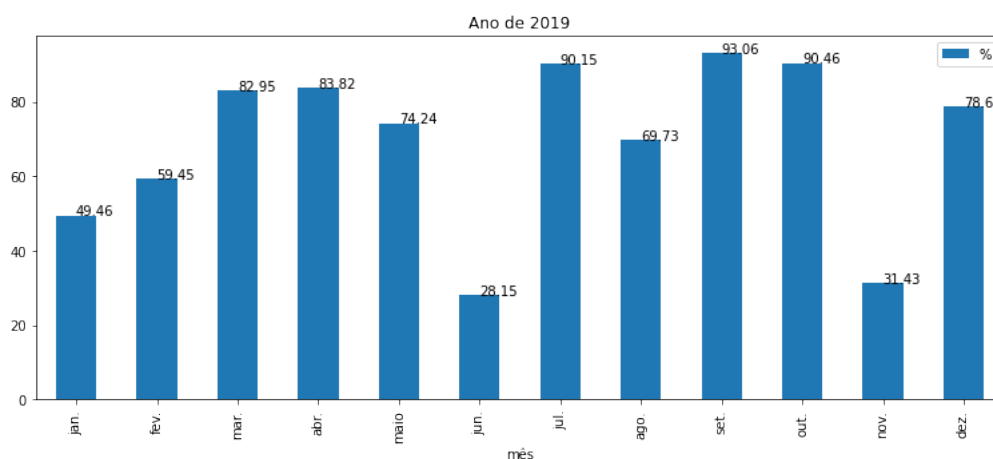


Figura 11 – Frequência dos tópicos predominantes de cada mês de 2019.

### 6.3 2020

Na Tabela 6, são apresentados os tópicos mais frequentes com as suas dez palavras mais importantes e a rotulação de cada mês do ano de 2020. Perceba que esse ano iniciou e terminou com tema sobre *Genocídio*. Ao longo do ano, o assunto *Genocídio* aparece como mais

frequente nos meses de janeiro, fevereiro, julho, setembro e dezembro (conforme a Tabela 6). O Apêndice C apresenta todos os cinco tópicos de cada mês.

Mês	Palavras	Assunto
jan.	general, massacre, missao, filme, governo, general_helene, filme_sobre, terror, precisar, haitiana	genocídio
fev.	lula, helene, ficar, general_helene, criminoso, povo, brasileiro, militar, deixar, gente	genocídio
mar.	haitiano, presidente, bolsonaro, brasil, brasileiro, mandar, ficar, governo, bolsonaro_acabar, sair	Intromissão política
abr.	haitiano, presidente, bolsonaro, acabar, governo, bolsonaro_acabar, vir_futuro, futuro, brasileiro, razao	Intromissão política
maio	haiti, haitiano, presidente, bolsonaro, governo, brasileiro, povo, lembrar, revolucao, tempo	Intromissão política
jun.	haiti, brasil, republica_dominicana, panama, colombia_equador, trinidad_tobago, pais, suriname_trinidad, banco_mundial, grupo	países do Banco Mundial
jul.	haiti, ficar, eua, brasileiro, grande, povo, governo, general, matar, ajudar	genocídio
ago.	haiti, brasil, cuba, pessoa, mundo, eua, gente, republica_dominicana, grande, senhor	tempestade Laura
set.	haiti, brasil, negro, brasileiro, gente, matar, grande, militar, pessoa, pobre	genocídio
out.	haiti, passar, eua, deixar, ficar, povo, olhar, antes, tornar, acontecer	imigração
nov.	haiti, brasil, eua, brasileiro, mundo, chegar, pais, pessoa, venezuela, pensar	imigração
dez.	haiti, brasil, gente, general, governo, helene, bolsonaro, pais, negro, venezuela	genocídio

Tabela 6 – Tópico predominante de cada mês de 2020.

O assunto *Genocídio* dominou os meses de janeiro, fevereiro, julho, setembro e dezembro. Refere-se a operação de pacificação realizada na maior favela da capital haitiana, Porto Príncipe, conhecida como *Cité Soleil*. Essa operação foi realizada na madrugada de 6 de julho de 2005 pelas tropas da Missão de Estabilização da ONU no Haiti (MINUSTAH). Segundo os testemunhas, cerca de 300 homens fortemente armados invadiram o bairro e assassinaram 63 pessoas, deixando outras 30 feridas. Sendo que, essas vítimas eram majoritariamente mulheres e crianças inocentes. A missão era sob o comando do general brasileiro Augusto Heleno, ministro-chefe do Gabinete de Segurança Institucional (GSI), responsável pela inteligência do governo de Jair Bolsonaro (PSL). Pode-se observar o extrato de um *tweet* sobre o tema: "*@bolsolulamaster Véio Heleno aloprou no Haiti. Responsável pela morte de mulheres, crianças e idosos. É GENOCIDA.*".

Os meses de março, abril e maio foram destacados pelo assunto *Intromissão política*. A discussão dos internautas desses meses referenciou a cena em que um jovem haitiano se aproximou da entrada do Palácio da Alvorada e falou o seguinte: "*Bolsonaro, acabou. Você está recebendo mensagem no celular. Todo brasileiro está recebendo mensagem no seu celular. Você não é presidente mais. Você não é presidente mais*", quando Bolsonaro se preparava para falar com seus apoiadores. Observa-se o extrato de *tweet* referente a esse tema: "*@jairbolsonaro Vim aqui pra mandar voce se fuder, mas vi que já tem muita gente fazendo isso, nem vai perder meu tempo com uma pessoa medíocre como você, o haitiano esta certinho, você não é presidente.*".

Nos meses de junho e agosto, a discussão era sobre *Países do Banco Mundial* e a *Tempestade Laura*, respectivamente. Em junho, o Partido dos Trabalhadores (PT) repudiou expressamente a indicação de Abraham Weintraub pelo governo brasileiro para ocupar o cargo de diretor-executivo no Banco Mundial. Essa indicação foi feita sem nenhuma consulta prévia dos demais países do grupo do Brasil no Banco Mundial, sendo que o Haiti é membro desse grupo. Já que, no mês de julho, os comentários referenciam a passagem da tempestade Laura, ventos de 120 km/h, que passou por Cuba mas antes disso já havia deixado pelo menos mais de 24 pessoas mortas no Haiti e na República Dominicana.

Como no ano de 2018, segundo a Tabela 4, o tema *Imigração* apareceu novamente em 2020 e dominou os meses de outubro e novembro. Nesses períodos, os comentários

falam, principalmente, sobre a imigração haitiana e venezuelana no Brasil e a imigração dos países latino-americanos nos Estados Unidos da América. Pode-se observar o extrato de *tweet* referente a esse tema: "*mesmo sendo na perifa, eu atendo muito imigrante aqui. quase todos são venezuelanos, colombianos, bolivianos e haitianos (os mais numerosos), tem até um romeno. cuiabá recebe MUITO imigrante latino-americano, o centro é cheio de ambulante haitiano, cheio mesmo*".

Na Figura 12, é mostrada a porcentagem do tópico mais frequente com relação aos outros tópicos de cada mês do ano durante o ano de 2020. Pode-se observar que oito de 12 meses apresentaram um valor maior do que 50%. Isto é, mais do que 50% dos *tweets* estão associados ao tópico mais frequente desses meses. Como exemplo, os meses de março e maio, representados pelo tópico *Intromissão política*, tiveram valores de 97,77% e 98,04%, nessa ordem.

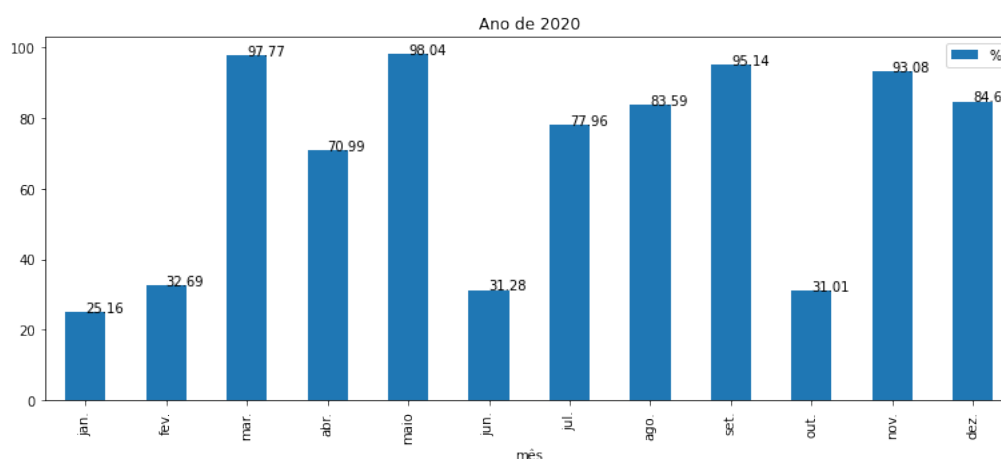


Figura 12 – Frequência dos tópicos predominantes de cada mês de 2020.

## 6.4 2021

No ano de 2021, a discussão dos internautas iniciou com tema sobre a *Independência haitiana* e terminou com assuntos sobre *Lembrança do atrevimento haitiano de 2020*. A Tabela 7 apresenta os tópicos mais frequentes de cada mês do ano com as suas dez palavras mais importantes e a rotulação de cada tópico. Neste ano, o assunto *Imigração* teve mais destaque, pois apareceu nos meses de junho, outubro e novembro. O Apêndice D apresenta todos os cinco tópicos de cada mês.

Nos meses de janeiro, fevereiro, março, abril, maio e dezembro, os *tweets* focavam mais sobre *Independência haitiana*, *Violação constitucional*, *Interferência política*, *Intervenção militar*, *Revolução haitiana* e *Lembrança do atrevimento haitiano de 2020*. Em janeiro, os comentários referem a independência do Haiti de 01 de janeiro de 1804, sendo a primeira nação negra do mundo. O mês de fevereiro foi dominado pela decisão do presidente Jovenel Moïse de se manter no poder como presidente depois de 7 de fevereiro, data em que acaba o seu mandato



Mês	Palavras	Assunto
jan.	haiti, brasileiro, grande, povo, independencia, ficar, primeiro, cuba, militar, governo	independência haitiana
fev.	haiti, brasil, governo, presidente, eua, ficar, contra, bolsonaro, grande, tentar	violação constitucional
mar.	haiti, brasil, governo, brasileiro, ficar, eua, grande, mundo, cuba, venezuela	interferência política
abr.	haiti, brasil, brasileiro, ficar, povo, país, eua, cuba, missao, deixar	intervenção militar
maio	haiti, revolucao, grande, povo, ajudar, general, passar, matar, pensar, historia	revolução haitiana
jun.	haiti, brasil, contra, eua, grande, brasileiro, gente, revolucao, povo, chegar	imigração e xenofobia
jul.	presidente, brasil, assassinar, jovenel_moise, brasileiro, matar_presidente, politico, bolsonaro, casa, matar	assassinato do presidente
ago.	terremoto, morto, passar, sofrer, grande, situacao, numero_morto, taliba, noticia, tristeza	terremoto
set.	haiti, presidente, ajudar, terremoto, grande, pessoa, gente, venezuela, deixar, passar	assassinato do presidente e terremoto
out.	haiti, brasil, eua, governo, pessoa, presidente, venezuela, americano, mundo, povo	imigração e refúgio
nov.	haiti, brasil, brasileiro, grande, mundo, ficar, povo, lula, negro, país	imigração e refúgio
dez.	haitiano, brasil, lula, brasileiro, governo, presidente, mundo, cuba, revolucao, acontecer	lembança do atrevimento haitiano de 2020

Tabela 7 – Tópico predominante de cada mês de 2021.

conforme a constituição haitiana. Isso gerou uma crise onde manifestantes têm saído às ruas em várias cidades do país. O desfecho dessa crise teve boa parte dos Estados Unidos, a potência mais influente sobre o Haiti. Em março, os comentários eram sobre o governo de Joe Biden, presidente dos Estados Unidos das Américas. Os internautas acham que esse governo não se atreveria a se intrometer caso houvesse tentativa de golpe no Brasil, como já tem feito em países como o Haiti e a Venezuela.

Como nos anos de 2018 e 2019, o tema Intervenção militar voltou a aparecer no ano de 2021 e dominou o mês de abril. Fez referência à intervenção do exército brasileiro que atuou no Haiti. O mês de maio foi marcado por comentários sobre a revolução haitiana que conduziu à independência do Haiti em 1804. E o mês de dezembro foi marcado pelos apoiadores do Lula comentando sobre o atrevimento do imigrante haitiano contra Bolsonaro discutido nos meses de março, abril e maio de 2020. Observa-se um *tweet* de exemplo: "*@BolsonaroSP É melhor você vestir o boné "Lula Presidente". Como disse um sábio haitiano, em Brasília: "você não é presidente mais". "seu governo acabou!". Tá lembrado? A onda do PSL 2018 tá virando marolinha. Vai ter que rebolar muito pra encontrar a tal reeleição."*

Junho, outubro e novembro foram notados por assuntos sobre *Imigração e Xenofobia* e *Imigração e refúgio*. Referem-se especialmente à imigração haitiana e venezuelana no Brasil e da imigração dos países latino-americanos nos Estados Unidos. Em junho, por exemplo, os assuntos relatam situações xenofóbicas que os imigrantes venezuelanos e haitianos têm passado no Brasil. Pode-se observar um *tweet* de exemplo sobre o tema: "*@JefinhoMenes Revoltados online, conseguiram reviver esse bosta kkkkkkk Eu só lembro dele no cqc sendo abordado por ter sido racista e xenofobico com um imigrante haitiano e ele só respondendo que era o Foro de São Paulo o culpado"*. E nos demais meses, a discussão era mais em torno da imigração ilegal do povo latino-americano nos Estados Unidos, inclusive os haitianos.

Os temas *Assassinato do presidente* e *Terremoto* dominaram os meses de julho, agosto e setembro. O mês de julho foi marcado pelo assassinato do presidente Jovenel Moïse no dia 7 de julho, na sua residência privada por indivíduos armados não identificados. Em agosto, os *tweets* focaram mais no terremoto de magnitude 7,2 que atingiu o Sul do Haiti, em 14 de agosto de 2021 e que deixou pelo menos 227 mortes. Por fim, em setembro, a discussão abrangia comentários tanto sobre o assassinato do presidente como sobre o terremoto de 14 de agosto.

Note-se o extrato de um *tweet* sobre os temas: "Em menos de 100 dias, o Haiti passou por um magnicídio, um terremoto e um furacão. Tudo isso contribuiu para o aumento do poder de gangues. 4/6 <https://t.co/cs1GAqLoop>".

Como nos anos anteriores, os *tweets* estão bem concentrados também nesse ano de 2021. Pode-se notar pela Figura 13, a porcentagem do tópico mais frequente de cada mês do ano. Dentre os 12 meses, 11 apresentam porcentagens acima de 50%. Em outras palavras, mais de 50% dos *tweets* estão associados a um tópico específico em cada um desses 11 meses. Como exemplo, os meses de outubro e novembro representados pelo assunto *Imigração e refúgio* apresentam, respectivamente, valores de 93.84% e 94.41%.

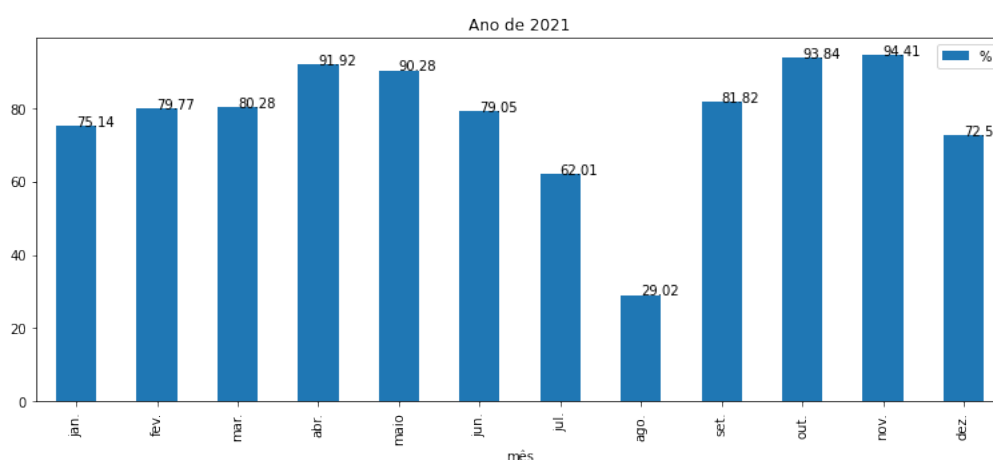


Figura 13 – Frequência dos tópicos predominantes de cada mês de 2021.

## 6.5 2022

O ano de 2022 compreende apenas *tweets* de meses de janeiro a maio. A Tabela 8 mostra os tópicos desses cinco meses de 2022 com as dez palavras com mais probabilidades de aparição e a rotulação dos tópicos. Esses períodos iniciaram e terminaram com tema sobre *Comparação sociopolítico-econômica*. Durante esses cinco meses, o tema *Racismo* aparece como mais frequente nos meses de fevereiro, março e abril. O Apêndice E apresenta todos os cinco tópicos de cada mês.

Mês	Palavras	Assunto
jan.	haiti, brasil, eua, presidente, ficar, pessoa, mundo, terremoto, governo, independencia	comparação sociopolítico-econômica
fev.	brasil, eua, cuba, guerra, mundo, pais, venezuela, lembrar, congo, ucrania	racismo
mar.	haiti, brasil, guerra, brasileiro, eua, cuba, ucrania, pessoa, grande, gente	racismo
abr.	haiti, brasileiro, ficar, pessoa, povo, coisa, passar, venezuela, conseguir, servir	racismo
maio	brasil, venezuela, bolsonaro, cuba, pais, mau, franca, gente, grande, governo	comparação sociopolítico-econômica

Tabela 8 – Tópico predominante de cada mês de 2022.

Os meses de janeiro e de maio foram notados por *tweets* sobre *Comparação sociopolítico-econômico*. As comparações de janeiro eram sobre a preocupação dos internautas, criticando

o governo do presidente Bolsonaro que está dirigindo o Brasil a caminho do Haiti. Observa-se um *tweet* de exemplo: "*@Rodrigo\_A\_Tw @Dessler\_ @Gui\_Ferrari\_ @Peruzzo\_8 @taoquei1 Cara Bolsonaro é embusteiro o pior presidente que Brasil já teve se ficar por mais 4 anos Brasil vira Haiti há mas ele é liberal nunca foi mas prega a liberdade só em frente as câmeras nos bastidores exonera e transfere quem discorda dele seus filhos nenhum projeto é só mamata*". Em maio, foram comentários sobre a conversa franca do presidente Jair Bolsonaro na imprensa, falando sobre o seu governo, respondendo às críticas que tem recebido dos opositores. Nestas falas, o presidente também fez comparações a países como Venezuela e Cuba.

Em fevereiro, março e abril, foram comentários referentes a imigração ucraniana causada pela guerra entre a Rússia e a Ucrânia. Percebeu-se lamentação por refugiados ucranianos até mesmo de pessoas que humilham ou não aceitam a imigração de pessoas como do Haiti, da Venezuela ou de outros países pobres. Na cidade de Maringá do Estado de Paraná (PR), em abril de 2022, os refugiados ucranianos são recebidos por festa realizada pela prefeitura da cidade, enquanto imigrantes de países pobres são maltratados. Observa-se o extrato de um *tweet* de exemplo: "*festinha pra receber ucraniano em maringá, enquanto venezuelanos e haitianos são tratados que nem lixo aqui. nunca vi uma ação de integração por parte da prefeitura, eles colocaram um haitiano como secretário em algum lugar*".

Como nos anos de 2018, 2019, 2020 e 2021, nesses cinco meses de 2022, os *tweets* também são bem centralizados para um determinado assunto. Pode-se perceber pela Figura 14, que em três dos cinco meses, mais de 50% dos *tweets* estão relacionados a um tópico. Por exemplo, os meses de janeiro e março, dominados pelos assuntos *Comparação sociopolítico-econômica* e *Racismo*, apresentam porcentagens de 92.35 e 95.54, respectivamente.

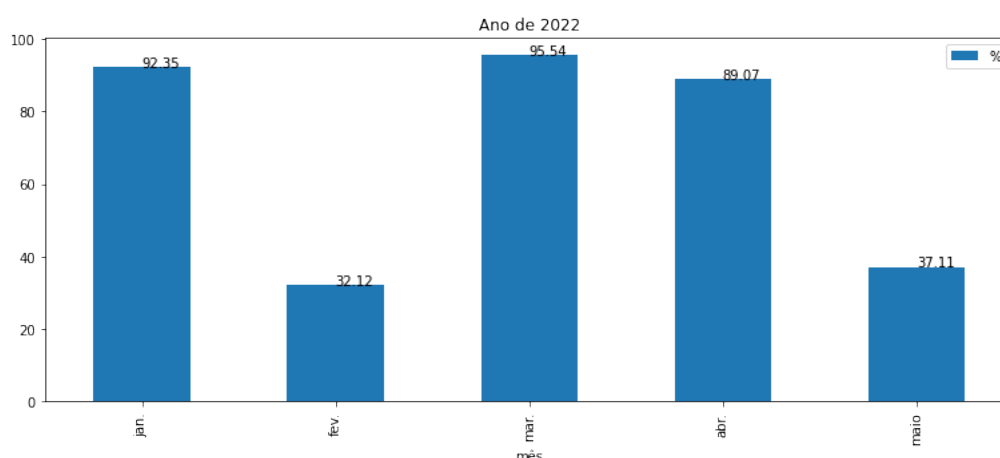


Figura 14 – Frequência dos tópicos predominantes dos cinco meses de 2022.

Neste capítulo, foram apresentados os resultados e discussões deste estudo. Mediante essas discussões, pode-se perceber que os tópicos obtidos são condizentes com a realidade. Assim dizendo, esses tópicos retratam fatos ou eventos que realmente aconteceram no Haiti ao longo do intervalo de janeiro de 2018 a maio de 2022, ou até mesmo de períodos precedentes a esse espaço de tempo, mas que foram discutidos durante esse intervalo no Twitter. Relatam,

também, fatos que tiveram relações com o povo haitiano, seja pelo que esse povo tem sofrido no mundo por meio de falas. Da mesma forma, mostram relações existentes entre acontecimentos no exterior com os do Haiti através de comparações.

## 7 CONCLUSÃO

Neste estudo, foi proposta uma análise exploratória das postagens mais discutidas sobre o Haiti no Twitter. Essas postagens (*tweets*) foram extraídas em língua portuguesa entre janeiro de 2018 a maio de 2022. Para a extração dos tópicos (assuntos), usou-se Latent Dirichlet Allocation (LDA) como abordagem de modelagem de tópicos.

Com os diversos acontecimentos que têm ocorrido no Haiti ao longo do tempo, foi-se necessário identificar os temas mais abordados no Twitter. Para atingir os objetivos, os dados foram coletados na plataforma do Twitter nas datas definidas. A seguir, realizou-se processo de pré-processamento dos *tweets* em que as palavras indesejadas ou parte dos *tweets* foram removidas. Em adição, aplicou-se o modelo LDA para extrair os tópicos, dividindo os dados da coleção em fatias mensais, ou seja, foram geradas 53 subcoleções. Nesse processo de extração, definiu-se o melhor número de tópicos que seriam extraídos pelo uso de métrica de avaliação do modelo para cada mês. O melhor número de tópicos de toda a coleção foi aquele que apareceu em mais meses, sendo 5 esse número, denotado por  $K$ .

Outra etapa que foi fundamental neste trabalho é a rotulação dos tópicos. É um processo que consiste em definir temas para os tópicos obtidos. Tal definição foi baseada em conhecimentos da coleção ou na análise das dez palavras que apresentaram as maiores probabilidades no tópico. Essa análise foi realizada no site de notícias do *Google* ou nos *tweets* dos períodos de cada tópico. Como a coleção original foi separada em 53 subcoleções, foram gerados 265 tópicos ( $53 \times 5$ ). Rotular todos os tópicos manualmente é custoso, assim, optou-se por rotular apenas o tópico que está associado a mais documentos em cada mês. Os Apêndices A, B, C, D e E apresentam as palavras de todos os tópicos gerados para cada subcoleção.

Os resultados foram apresentados por meio de tabelas contendo as palavras geradas e gráficos que representam a porcentagem de cada tópico rotulado em relação aos outros de cada mês de cada ano. Mediante discussões desses resultados, percebe-se que os tópicos extraídos descrevem exatamente situações correspondentes ao Haiti. Portanto, os objetivos propostos neste estudo foram alcançados.

Assim, este trabalho pode servir como base de conhecimentos tanto para pesquisadores da área de ciências sociais, histórias e áreas conexas como para qualquer leitor. Como também, pode-se servir como base de pesquisas sobre um determinado assunto referente ao Haiti no mesmo período de tempo, mas em qualquer idioma. Ou pesquisas de diversos assuntos mas, em períodos diferentes. Como direções para futuros trabalhos pode-se citar: (i) rotular todos os tópicos e relacioná-los em cada fatia e (ii) analisar a evolução dos assuntos (tópicos) no tempo. Por exemplo, identificar como um assunto nasce ou deixa de ser discutido ao longo do tempo.

## REFERÊNCIAS

- ASGHARI, Mohsen; SIERRA-SOSA, Daniel; ELMAGHRABY, Adel. Trends on health in social media: Analysis using twitter topic modeling. In: IEEE. 2018 IEEE international symposium on signal processing and information technology (ISSPIT). [S.l.: s.n.], 2018. p. 558–563.
- BBC. **Os 10 terremotos mais potentes e com maior número de mortos da história da América Latina**. 2017. Disponível em: <<https://www.bbc.com/portuguese/internacional-41380495>>. Acesso em: 20 jan. 2022.
- BIASSETTE, Gilles. **Assassinato do presidente Jovenel Moise. La Croix**. 2021. Disponível em: <<https://www.la-croix.com/Monde/Haiti-president-Jovenel-Moise-assassine-commando-2021-07-07-1201165212>>. Acesso em: 19 jan. 2022.
- BLEI, David M. Probabilistic topic models. **Communications of the ACM**, Association for Computing Machinery (ACM), v. 55, n. 4, p. 77, abr. 2012. DOI: 10.1145/2133806.2133826. Disponível em: <<https://doi.org/10.1145/2133806.2133826>>.
- CHARLES, Jacqueline. **Dez anos após ser devastado por terremoto, Haiti ainda vive em cenário desolador**. 2020. Disponível em: <<https://about.twitter.com/content/dam/about-twitter/en/company/global-impact-2020.pdf>>. Acesso em: 19 jan. 2022.
- CHAUHAN, Uttam; SHAH, Apurva. Topic modeling using latent Dirichlet allocation: A survey. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 7, p. 1–35, 2021.
- CHEHAL, Dimple; GUPTA, Parul; GULATI, Payal. Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 12, n. 5, p. 5055–5070, 2021.
- DAVID M. BLEI ANDREW Y. NG, Michael I. Jordan. Latent Dirichlet Allocation, 2003.
- DUARTE, Denio; STÅHL, Niclas. Machine Learning: A Concise Overview. In: STUDIES in Big Data. [S.l.]: Springer International Publishing, set. 2018. p. 27–58. DOI: 10.1007/978-3-319-97556-6\_3. Disponível em: <[https://doi.org/10.1007/978-3-319-97556-6\\_3](https://doi.org/10.1007/978-3-319-97556-6_3)>.
- FRANCISCO, Wagner de Cerqueira e. **"Haiti - Aspectos Geográficos"; Brasil Escola**. Disponível em: <<https://brasilecola.uol.com.br/geografia/haitiaspectos-geograficos.htm>>. Acesso em: 30 mai. 2022.

- GUITARRARA, Paloma. "**Terremotos no Haiti**"; *Brasil Escola*. Disponível em: <<https://brasilecola.uol.com.br/geografia/o-terremoto-no-haiti.htm>>. Acesso em: 19 mai. 2022.
- HIDAYATULLAH, Ahmad Fathan et al. Twitter topic modeling on football news. In: IEEE. 2018 3rd International Conference on Computer and Communication Systems (ICCCS). [S.l.: s.n.], 2018. p. 467–471.
- HURBON, Laënnec. Le vodou et la révolution haïtienne. **Tumultes**, Éditions Kimé, v. 50, n. 1, p. 59–72, 2018.
- OH, Onook; KWON, Kyounghee Hazel; RAO, H Raghav. An exploration of social media in extreme events: Rumor theory and Twitter during the Haiti earthquake 2010, 2010.
- PEREIRA, Mariana. Análise exploratória de tweets utilizando modelagem de tópicos para textos curtos: caso Olimpíadas Rio 2016. Universidade Federal da Fronteira Sul, 2019.
- RÖDER, Michael; BOTH, Andreas; HINNEBURG, Alexander. Exploring the space of topic coherence measures. In: PROCEEDINGS of the eighth ACM international conference on Web search and data mining. [S.l.: s.n.], 2015. p. 399–408.
- ROHR, Linda; SQUIRES, Laura; PETERS, Adrienne. Examining the Use of Twitter in Online Classes: Can Twitter Improve Interaction and Engagement?. **Canadian Journal for the Scholarship of Teaching and Learning**, ERIC, v. 13, n. 1, n1, 2022.
- RTS. **Haiti: Breve história**. 2011. Disponível em: <<https://www.rts.ch/decouverte/dossiers/2010/haiti/2876706-histoire-de-haiti.html#:~:text=L'%C3%AEle%20d'Hispaniola%20a,auxquels%20les%20Espagnols%20les%20soumettent.>>. Acesso em: 19 jan. 2022.
- SILVA, Adeilson Souza da et al. Detectando comportamento automatizado nos tópicos de tendência do Twitter no Brasil. Universidade Federal do Amazonas, 2015.
- SOUZA, Rafaela. **Imigração haitiana no Brasil**. Disponível em: <<https://mundoeducacao.uol.com.br/geografia/imigracao-haitiana-no-brasil.htm>>. Acesso em: 19 jan. 2022.
- STEYVERS, Mark; GRIFFITHS, Tom. Probabilistic topic models. **Handbook of latent semantic analysis**, v. 427, n. 7, p. 424–440, 2007.
- TROUILLARD, Stéphanie. **Assassinato do presidente Jovenel Moise. France24**. 2021. Disponível em: <<https://www.france24.com/fr/info-en-continu/20210707-ha%C3%Afti-le-pr%C3%A9sident-jovenel-mo%C3%AFse-assassin%C3%A9-dans-la-nuit-dans-sa-r%C3%A9sidence-priv%C3%A9e-cabinet-du-premier-ministre>>. Acesso em: 19 jan. 2022.
- TWITTER. **Global impact report**. 2020. Disponível em: <<https://about.twitter.com/content/dam/about-twitter/en/company/global-impact-2020.pdf>>. Acesso em: 19 jan. 2022.

VERENHITACH, Gabriela Daou et al. A MINUSTAH e a política externa brasileira: motivações e consequências. Universidade Federal de Santa Maria, 2008.



## A APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2018

Tópicos	Palavras	Documentos associados
0	haiti, trump, brasil, eua, africa, pessoa, brasileiro, chamar, onu, terremoto	1842
1	haitiano, vida, lugar, conhecer, casa, amigo, entrar, ficar, branco, levar	251
2	pais, presidente, salvador, pais_africano, donald_trump, imigrante, merda, referir, merda, perguntar	166
3	trump_chamar, merda, buraco_merda, mundo, america_central, africa, outro_pais, verdade, pais, jornal	63
4	povo, gente, cuba, negro, morrer, viver, trabalhar, coisa, morar, chegar	25

Tabela 9 – Tópicos de janeiro de 2018.

Tópicos	Palavras	Documentos associados
0	exercito, militar, general, brasileiro, povo, rio_janeiro, trabalho, usar, coisa, sexual	439
1	rio, onu, presidente, comandante, missao_paz, ajudar, contra, forca_armada, intervencao, general_helena	373
2	brasil, oxfam, missao, escandalo_sexual, matar, guerra, mundo, caso, pessoa, chegar	322
3	haitiano, gente, venezuela, grande, problema, exercito_brasileiro, parecer, entender, acoes, chamar	261
4	ficar, mandar, eua, haitiana, comparar, pensar, tempo, tropa, merval_aqui, bolsonaro	114

Tabela 10 – Tópicos de fevereiro de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, rio, gente, exercito, brasileiro, militar, sirio, ajudar, venezuela	1000
1	haitiano, negro, chegar, pensar, trabalhar, entender, branco, video, imigrante, kaysar	159
2	terremoto, olhar, usar, viver, colocar, missao_paz, cuba, mexico, revolucao, forca_armada	14
3	onu, ficar, pessoa, casa, contra, refugiar, conhecer, pedir, estar, cidade	14
4	passar, haitiana, guerra, vida, amigo, mundo, homem, tempo, crianca, esperar	10

Tabela 11 – Tópicos de março de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, exercito, usar, militar, vida, missao_paz, terremoto, mandar, missao	408
1	haitiano, ficar, refugiar, kaysar, brasil, vender, venezuelano, branco, sirio, gente	335
2	haiti, povo, ajudar, brasil, venezuela, sirio, mundo, acontecer, africa, guerra	159
3	haiti, eua, gente, haitiana, lula, brasil, refugiar, receber, outro_pais, venezuela	119
4	haiti, onu, brasil, escola, aluno, brasileiro, morrer, vender_livro, pedir, grupo	106

Tabela 12 – Tópicos de abril de 2018.

Tópicos	Palavras	Documentos associados
0	contra, messi, selecao, jogo, grande, cuba, argentino, amistoso, fazer_gol, hat_trick	486
1	haitiano, ficar, venezuela, pessoa, capitalismo, coisa, negro, pais, refugiar, viver	440
2	argentina, copa, jogar, passar, gol, golear, primeiro, pegar, mundo, levar	388
3	brasil, gente, bombonera, antes, terremoto, ontem, amistoso_contra, militar, exercito, vida	343
4	brasileiro, povo, ganhar, general, senhor, precisar, ajudar, deixar, missao, haitiana	192

Tabela 13 – Tópicos de maio de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, brasileiro, selecao, povo, ficar, ajudar, futebol, pessoa, venezuela	1389
1	copa, jogo, contra, argentina, jogar, time, messi, alemanha, jogador, selecoes	106
2	haitiano, passar, acabar, tentar, entrar, amigo, parecer, revolucao, olhar, vender	46
3	gente, grande, lembrar, lugar, video, viver, casa, negro, virar, galerar	20
4	mundo, primeiro, cuba, tempo, precisar, jamaica, mexico, panamar, gostar_video, sofrer	14

Tabela 14 – Tópicos de junho de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, gente, brasileiro, jamaica, ficar, selecao, chorar, pessoa, mundo, pais	1797
1	haitiano, the_voice, cantar, chegar, imigrante, amigo, pegar, filho, musica, acabar	131
2	brasil, governo, coisa, rua, haitiana, jogo, venezuela, povo, terremoto, protesto	33
3	primeiro_ministro, povo, contra, crise, trabalho, onu, ajudar, gostar, time, premie	6
4	cuba, grande, ganhar, passar, mandar, estar, amor, coracao, militar, conseguir	2

Tabela 15 – Tópicos de julho de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, militar, brasileiro, ficar, chegar, bolsonaro, missao, mundo, grande	1504
1	haitiano, gente, coisa, passar, perguntar, amigo, haitiana, parar, revolucao, rua	230
2	exercito, pacificar, mandar, exercito_brasileiro, acontecer, bolsonaro, presidente, virar, primeiro, arma	42
3	onu, ajudar, parecer, trabalho, precisar, acabar, curitiba, luciano_huck, guerra, morrer	39
4	contra, venezuelano, pessoa, venezuela, deixar, mulher, povo, trabalhar, ganhar, sirio	21

Tabela 16 – Tópicos de agosto de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasileiro, passar, pessoa, missao_paz, contra, governo, mundo, missao, militar	843
1	haitiano, gente, voltar, venezuelano, rua, branco, por, filha, entender, mae_japones	255
2	brasil, venezuela, haiti, cuba, mau, grande, pensar, america_latina, desemprego_grande, presidente	169
3	haitiano, haitiana, trabalhar, chamar, japonesa, japonês, serena, ganhar, naomir_osaka, vencer	105
4	onu, chegar, filho, igual, casa, conhecer, levar, povo, morar, situacao	62

Tabela 17 – Tópicos de setembro de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, povo, cuba, brasileiro, passar, militar, presidente, paz, conhecer, guerra	558
1	haiti, brasil, venezuela, onu, mundo, bolsonaro, eua, igual, pobre, chegar	301
2	haitiano, ficar, brasileiro, centro, amigo, passar, conseguir, voltar, entender, imigrante	286
3	haiti, grande, missao_paz, bolsonaro, capitalista, haitiana, governo, coisa, parecer, exercito	237
4	haiti, terremoto, pessoa, ajudar, gente, negro, matar, acabar, vida, viver	218

Tabela 18 – Tópicos de outubro de 2018.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, grande, ajudar, mundo, contra, eua, usar, coisa, deixar	1312
1	onu, militar, governo, general, brasileiro, missao, congo, venezuela, exercito, povo	346
2	haitiano, pessoa, negro, haitiana, bolsonaro, chegar, enem, revolucao, vida, igual	232
3	cuba, medico, pais, cubano, pagar, antes, medico_cubano, receber, programa, precisar	158
4	haitiano, gente, conhecer, guerra, chamar, olhar, ficar, brasileiro, mostrar, perguntar	97

Tabela 19 – Tópicos de novembro de 2018.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, militar, brasileiro, imigrante, bolsonaro, haitiana_cega, governo, povo, aprovar_oab	1191
<b>1</b>	haitiano, ficar, gente, acabar, grande, pessoa, venezuelano, casa, sair, passar	357
<b>2</b>	cuba, rua, voltar, pais, eua, mundo, paz, servir, apenas, virar	21
<b>3</b>	ajudar, haitiana, negro, trabalhar, estar, branco, mae, publico, porto_alegre, natal	21
<b>4</b>	chegar, general, filho, vida, familia, levar, refugiar, contar, aprender, morar	3

Tabela 20 – Tópicos de dezembro de 2018.

## B APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2019

Tópicos	Palavras	Documentos associados
0	haiti, terremoto, ajudar, brasileiro, exercito_brasileiro, militar, brumadinho, eua, atuar, precisar	1234
1	haitiano, parecer, policial, conseguir, casa, morrer, contar, venezuelano, gente, pagar	472
2	exercito, haiti, brasil, ajudar, militar, israel, governo, humanitario, vira, ajuda	298
3	brasil, haiti, ficar, onu, mandar, lembrar, general, usar, presidente, gostar_video	247
4	haiti, grande, venezuela, mundo, gente, pessoa, primeiro, revolucao, trabalho, pensar	244

Tabela 21 – Tópicos de janeiro de 2019.

Tópicos	Palavras	Documentos associados
0	eua, petroleo, venezuela, humanitario, ajuda_humanitaria, ajuda_humanitaria, mundo, trump, mandar, pais	1991
1	brasil, ajudar, ficar, onu, passar, militar, senador, terremoto, problema, enviar	1044
2	povo, governo, deixar, situacao, contra, usar, mau, guerra, exercito_brasileiro, venezuelano	184
3	haitiano, acontecer, protesto, senhor, matar, chamar, negro, tambem, contar, sair	76
4	brasileiro, presidente, gente, ajuda, pessoa, coisa, morrer, estar, parecer, chegar	54

Tabela 22 – Tópicos de fevereiro de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, venezuela, eua, ajudar, governo, petroleo, cuba, grande, mundo	1688
1	haitiano, ficar, pessoa, amigo, coisa, gente, pensar, perguntar, chamar, venezuelano	291
2	militar, brasileiro, onu, presidente, voltar, sair, exercito, mau, corrupcao, missao	38
3	trabalhar, general, negro, trabalho, acontecer, branco, conhecer, familia, entrar, primeiro	12
4	povo, passar, chegar, contra, problema, politico, acabar, mostrar, dinheiro, venezuela	6

Tabela 23 – Tópicos de março de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, pessoa, gente, ajudar, ficar, mundo, brasileiro, mocambique, venezuela, coisa	1425
1	haitiano, trabalho, pedir, caso, mulher, cada, centro, tentar, gostar, casa	199
2	brasil, exercito, militar, missao, general, bolsonaro, viver, precisar, lembrar, missao_paz	35
3	pensar, passar, grande, vida, tempo, pobre, guerra, chamar, deixar, grupo	22
4	acabar, perguntar, entender, haitiana, revolucao, trabalhar, comecar, branco, primeiro, preciso	19

Tabela 24 – Tópicos de abril de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, brasileiro, venezuela, ficar, jovem, ajudar, grande, gente, governo	1363
1	haitiano, chegar, chamar, casa, rua, perguntar, lado, gente, dinheiro, comprar	240
2	militar, congo, onu, general, presidente, bolsonaro, missao, missao_paz, santos_cruz, minustah	108
3	eua, povo, mandar, petroleo, caso, pedir, usar, parecer, acontecer, free_market	69
4	mundo, revolucao, pensar, pessoa, negro, vida, conhecer, primeiro, haitiana, trabalho	56

Tabela 25 – Tópicos de maio de 2019.

Tópicos	Palavras	Documentos associados
0	brasil, costa_rica, copa_ouro, grupo, jogo, eua, nicaragua, copa_america, curacao, contra	684
1	haitiano, heleno, ficar, passar, coisa, povo, deixar, matar, mulher, acontecer	652
2	onu, brasileiro, contra, lula, presidente, general, governo, ganhar, jogar, general_heleno	426
3	general, lula, selecao, missao, exercito, chile, pedir, lembrar, tempo, contar	347
4	pessoa, massacre, general_heleno, usar, ajudar, mandar, comandar, pegar, acabar, gente	321

Tabela 26 – Tópicos de junho de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, mexico, virar, jogo, eua, copa_ouro, ouro, grande, perder	1868
1	haitiano, acabar, amigo, revolucao, mae, casa, hora, familia, centro, pensar	90
2	passar, bolsonaro, militar, gente, governo, general_heleno, coisa, vergonha, mau, haitiana	53
3	brasileiro, general, ganhar, povo, pessoa, pais, deixar, venezuela, heleno, apenas, jornal	41
4	contra, prorrogacao, selecao, costa_rica, negro, minuto, penalti, gostar, campanha, mexico	20

Tabela 27 – Tópicos de julho de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, brasileiro, chamar, presidente, mundo, governo, pobre, povo, general	1569
1	haitiano, grande, ficar, gente, eua, revolucao, chegar, negro, voltar, trabalhar	516
2	deixar, franca, ajudar, colonia, macron, franco, francês, africa, mandar, dinheiro	106
3	amazonia, franco, cuidar, casa, viver, morar, sair, colonio, menos, congo	34
4	passar, pessoa, colonia_francês, haitiana, ouvir, ontem, caso, historia, politico, macron	25

Tabela 28 – Tópicos de agosto de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, general, brasileiro, pessoa, povo, deixar, contra, mundo, pais, coisa	1919
1	brasil, ficar, onu, matar, grande, militar, curacao, presidente, heleno, exercito	120
2	haitiano, acabar, olhar, esquecer, contar, perguntar, escola, verdade, trabalhar, povo	18
3	gente, negro, revolucao, pensar, branco, lugar, amigo, explicar, gostar, tentar	5
4	senhor, passar, missao, vida, chamar, pedir, haitiana, tirar, verdade, lula	0

Tabela 29 – Tópicos de setembro de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, equador, protesto, presidente, mundo, chile, gente, rua, hong_kong	2087
1	haitiano, revolucao, amigo, negro, chamar, imigrante, casa, branco, acabar, conhecer	102
2	povo, governo, contra, argentina, hondura, lutar, pais, peru, mexico, independencia	68
3	pessoa, jogo, viver, primeiro, tempo, comecar, coreia_sul, lembrar, politico, franco	25
4	ajudar, grupo, senhor, matar, lula, coisa, perder, voltar, pobre, estadio_serrinha	25

Tabela 30 – Tópicos de outubro de 2019.

Tópicos	Palavras	Documentos associados
0	haiti, heleno, contra, eua, negro, cuba, mundo, pensar, pais, pobre	877
1	haiti, brasil, militar, equador, chile, bolivia, franco, haitiana, lembrar, venezuela	539
2	haitiano, revolucao, chegar, perguntar, mulher, sobre_revolucao, ouvir, branco, passar, ficar	538
3	haiti, general, povo, brasileiro, lula, governo, mandar, massacre, sair, pessoa	521
4	haiti, coisa, chamar, gente, grande, acontecer, olhar, esperar, explicar, acabar	315

Tabela 31 – Tópicos de novembro de 2019.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, onu, soldado, filho, mulher, missao, ajudar, pessoa, vida	1456
<b>1</b>	haitiano, ficar, gente, passar, revolucao, trabalhar, conseguir, entender, chamar, comecar	315
<b>2</b>	brasileiro, militar, crianca, missaoa_paz, minustah, abandonar, acusar_estupro, tempo, menina, filho	30
<b>3</b>	governo, estar, deixar, general, haitiana, politico, capitalismo, existir, lula, lugar	26
<b>4</b>	povo, grande, lembrar, levar, amigo, esperar, precisar, contar, acabar, matar	24

Tabela 32 – Tópicos de dezembro de 2019.

## C APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2020

Tópicos	Palavras	Documentos associados
0	general, massacre, missao, filme, governo, general_helena, filme_sobre, terror, precisar, haitiana	506
1	haitiano, revolucao, gente, povo, chamar, mandar, chegar, presidente, negro, usar	484
2	brasil, ficar, grande, pessoa, eua, mundo, passar, janeiro, primeiro, coisa	397
3	brasileiro, terremoto, militar, deixar, filho, ajudar, morrer, olhar, sair, crianca	356
4	helena, onu, senhor, matar, contar, lembrar, guerra, contra, exercito, historia	268

Tabela 33 – Tópicos de janeiro de 2020.

Tópicos	Palavras	Documentos associados
0	lula, helena, ficar, general_helena, criminoso, povo, brasileiro, militar, deixar, gente	680
1	general, matar, massacre, receber, papa, assassino, senhor, crianca, miliciano, exercito	606
2	haitiano, passar, ligar, esquecer, entender, acabar, hora, estar, ajudar, levar	370
3	baile, pessoa, grande, chegar, esperar, coisa, pegar, mundo, contar, caso	222
4	brasil, governo, onu, pensar, lembrar, missao, morte, crime, casa, comandar	202

Tabela 34 – Tópicos de fevereiro de 2020.

Tópicos	Palavras	Documentos associados
0	haitiano, presidente, bolsonaro, brasil, brasileiro, mandar, ficar, governo, bolsonaro_acabar, sair	18634
1	haiti, volta, voltar_pro, volta_pro, mundo, estar, ajudar, gostar, lugar, levar	387
2	povo, precisar, verdade, coragem, estrangeiro, matar, cair, tempo, parecer, parabens	17
3	voltar, gente, pensar, deve_ser, pagar, merda, passar, representar, deportar, tomar	17
4	acabar, bozo, coisa, chegar, imigrante, cidadao, direito, certo, olhar, governo	4

Tabela 35 – Tópicos de março de 2020.

Tópicos	Palavras	Documentos associados
0	haitiano, presidente, bolsonaro, acabar, governo, bolsonaro_acabar, vir_futuro, futuro, brasileiro, razao	3803
1	haiti, brasil, ficar, passar, povo, militar, terremoto, pensar, deixar, coisa	1501
2	gente, grande, pessoa, caso, general, chegar, primeiro, fim, faltar, politico	33
3	verdade, mundo, sair, video, amigo, moro, alguem, olhar, antes, povo	15
4	certo, acabar_bolsonaro, avisar, mandar, casa, conseguir, militar, presidente, ficar, haitiano	5

Tabela 36 – Tópicos de abril de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, haitiano, presidente, bolsonaro, governo, brasileiro, povo, lembrar, revolucao, tempo	3894
<b>1</b>	heleno, general_heleno, onu, matar, missao, passar, massacre, lula, comandar, explicar	36
<b>2</b>	brasil, general, militar, grande, contra, exercito, genocida, mundo, ajudar, levar	34
<b>3</b>	acabar, gente, pensar, amigo, coisa, chamar, pais, pobre, cuba, acontecer	6
<b>4</b>	ficar, pessoa, voltar, conhecer, estar, vergonha, eua, vida, olhar, contar	2

Tabela 37 – Tópicos de maio de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, republica_dominicana, panama, colombia_equador, trinidad_tobago, pais, suriname_trinidad, banco_mundial, grupo	1114
<b>1</b>	haitiano, presidente, revolucao, negro, bolsonaro, branco, lembrar, primeiro, bolsonaro_acabar, cair	994
<b>2</b>	haiti, general, militar, brasileiro, heleno, senhor, onu, pensar, gente, cuba	687
<b>3</b>	haiti, mundo, eua, ajudar, pessoa, tempo, haitiana, independencia, amigo, morrer	407
<b>4</b>	haitiano, acabar, governo, sair, mandar, chegar, caso, precisar, filipinas_suriname, esperar	359

Tabela 38 – Tópicos de junho de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, ficar, eua, brasileiro, grande, povo, governo, general, matar, ajudar	1592
<b>1</b>	haitiano, revolucao, haitiana, sobre_revolucao, historia, mulher, casa, vodu, vida, livro	330
<b>2</b>	brasil, gente, pensar, chamar, coisa, primeiro, chegar, antes, mostrar, americano	57
<b>3</b>	negro, branco, militar, pais, cuba, mundo, venezuela, pobre, exercito, lugar	50
<b>4</b>	acabar, bolsonaro, presidente, pessoa, passar, contra, olhar, conseguir, covid, virar	13

Tabela 39 – Tópicos de julho de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, cuba, pessoa, mundo, eua, gente, republica_dominicana, grande, senhor	2093
<b>1</b>	haitiano, sobre_revolucao, vodu, ler, amigo, agosto, historia, pegar, ontem, conseguir	305
<b>2</b>	pais, ajudar, grupo, brasil, tobago, independencia, suriname_trinidad, menos, colombia_equador, republica_dominicana	42
<b>3</b>	revolucao, chamar, haitiana, pensar, texto, coisa, gostar, america, livro, francês	35
<b>4</b>	negro, ficar, acabar, contra, branco, presidente, vida, perguntar, africano, general	29

Tabela 40 – Tópicos de agosto de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, negro, brasileiro, gente, matar, grande, militar, pessoa, pobre	2464
<b>1</b>	haitiano, revolucao, vida, branco, trabalho, voltar, existir, mulher, rua, parecer	64
<b>2</b>	viver, povo, historia, mundo, contar, cuba, exercito, usar, pais, historia	23
<b>3</b>	ficar, eua, chamar, presidente, independencia, primeiro, guerra, acabar, precisar, antes	21
<b>4</b>	general, governo, passar, filho, coisa, deixar, amigo, olhar, haitiana, conhecer	18

Tabela 41 – Tópicos de setembro de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, passar, eua, deixar, ficar, povo, olhar, antes, tornar, acontecer	627
<b>1</b>	haitiano, revolucao, chamar, lado, haitiana, trabalho, pensar, pedir, primeiro, casa	485
<b>2</b>	haiti, brasileiro, mau, heleno, general, onu, general_heleno, militar, jogar, crianca	370
<b>3</b>	haiti, brasil, venezuela, grande, governo, missao, coisa, viver, cuba, vida	331
<b>4</b>	haiti, mundo, hondura, pais, america_latina, branco, acabar, direito, conseguir, unico	209

Tabela 42 – Tópicos de outubro de 2020.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, eua, brasileiro, mundo, chegar, pais, pessoa, venezuela, pensar	1669
<b>1</b>	haitiano, revolucao, acabar, deixar, mae, trabalhar, mulher, pedir, trump, preto	88
<b>2</b>	negro, povo, branco, grande, matar, americano, usar, antes, amigo, escravo	18
<b>3</b>	gente, contra, guerra, existir, congo, casa, trabalho, tempo, politico, olhar	11
<b>4</b>	ficar, general, cuba, lembrar, bolsonaro, haitiana, exercito, igual, primeiro, precisar	7

Tabela 43 – Tópicos de novembro de 2020.



<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haiti, brasil, gente, general, governo, heleno, bolsonaro, pais, negro, venezuela	1651
<b>1</b>	haitiano, acabar, ficar, pessoa, chamar, pensar, pedir, mandar, mae, parecer	280
<b>2</b>	brasileiro, grande, eua, contra, mundo, lembrar, general_helena, matar, haitiana, baile	9
<b>3</b>	povo, cuba, presidente, deixar, trabalho, amigo, dinheiro, levar, rua, gostar	6
<b>4</b>	revolucao, ajudar, passar, gremio, entender, verdade, vida, jogar, primeiro, viver	4

Tabela 44 – Tópicos de dezembro de 2020.

## D APÊNDICE - TÓPICOS EXTRAÍDOS DO ANO DE 2021

Tópicos	Palavras	Documentos associados
0	haiti, brasileiro, grande, povo, independencia, ficar, primeiro, cuba, militar, governo	1862
1	haitiano, presidente, bolsonaro, gente, pessoa, acabar, chamar, voodoo, igual, gostar	377
2	revolucao, negro, terremoto, deixar, pensar, acontecer, chegar, olhar, coisa, haitiana	116
3	favor_ajudar, ajudar, favor, pequeno, brasil, salario_minimo, argentina, trazer_filha, brasileiro, filha	86
4	brasil, passar, lembrar, acabar, precisar, mundo, vida, sair, estar, pedir	37

Tabela 45 – Tópicos de janeiro de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, governo, presidente, eua, ficar, contra, bolsonaro, grande, tentar	2165
1	haitiano, povo, pessoa, mundo, imigrante, apoiar, rua, deixar, viver, equador	272
2	vencer, teste, querer_doar, milhao_teste, doar, covid, gente, quase_vencer, enviar, teste_covid	187
3	brasileiro, revolucao, general, militar, mandar, pensar, passar, tempo, precisar, populacao	48
4	estar, cuba, acabar, negro, coisa, golpe, exercito, sair, ajudar, chamar	42

Tabela 46 – Tópicos de fevereiro de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, governo, brasileiro, ficar, eua, grande, mundo, cuba, venezuela	2540
1	haitiano, presidente, bolsonaro, acabar, revolucao, lembrar, chamar, mulher, certo, pegar	530
2	gente, contra, pessoa, tempo, grupo, pais, haitiana, hondura, jogo, sair	58
3	povo, precisar, lula, mandar, acontecer, ajudar, mau, pedir, missao, apenas	28
4	militar, antes, coisa, virar, parecer, chegar, olhar, selecao, video, heleno	8

Tabela 47 – Tópicos de março de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, brasileiro, ficar, povo, pais, eua, cuba, missao, deixar	2310
1	haitiano, gente, revolucao, chegar, passar, lembrar, matar, entender, historia, parar	107
2	bolsonaro, governo, mundo, grande, pessoa, acabar, presidente, militar, contra, exercito	71
3	crianca_africana, ajudar, feliciano, negro, guerra, amigo, levar, africana, chamar, lugar	22
4	venezuela, africa, general_heleno, viver, general, crianca, pobre, conhecer, dinheiro, gente	3

Tabela 48 – Tópicos de abril de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, revolucao, grande, povo, ajudar, general, passar, matar, pensar, historia	1728
1	haitiano, acabar, gente, coisa, pessoa, tentar, sobre_revolucao, trabalho, venezuelano, situacao	92
2	brasil, ficar, eua, negro, lembrar, cuba, pais, venezuela, missao, precisar	78
3	presidente, governo, brasileiro, bolsonaro, rua, heleno, mandar, usar, chamar, chegar	11
4	mundo, exercito, contra, acontecer, pobre, militar, onu, deixar, perguntar, primeiro	5

Tabela 49 – Tópicos de maio de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, contra, eua, grande, brasileiro, gente, revolucao, povo, chegar	2701
1	haitiano, acabar, ficar, tempo, acaba_bolsonaro, video, futuro, vir_futuro, bolsonaro_acabar, governo	569
2	selecao, jogo, canado, passar, exercito, gol, canada, deixar, jogar, copa_america	67
3	goleiro, salvador, grupo, mundo, sao_cristovao, olhar, curacao, lula, pensar, eliminatória_concacaf	55
4	presidente, bolsonaro, pessoa, cair, negro, nicaragua, amigo, comecar, ouvir, pensar	25

Tabela 50 – Tópicos de junho de 2021.

Tópicos	Palavras	Documentos associados
0	presidente, brasil, assassinar, jovenel_moise, brasileiro, matar_presidente, politico, bolsonaro, casa, matar	7791
1	haitiano, cuba, eua, povo, mundo, contra, americano, grande, viver, revolucao	4143
2	militar, onu, ajudar, pensar, pais, coisa, entender, mandar, pedir, primeiro	258
3	assassinato_presidente, governo, acontecer, ficar, primeira_dama, venezuela, deixar, cia, estados_unidos, golpe	236
4	primeiro_ministro, gente, moise, passar, pessoa, prender, crise, policia, antes, mercenario	136

Tabela 51 – Tópicos de julho de 2021.

Tópicos	Palavras	Documentos associados
0	terremoto, morto, passar, sofrer, grande, situacao, numero_morto, taliba, noticia, tristeza	1572
1	haitiano, povo, brasileiro, presidente, governo, humanitario, bolsonaro, paz, contar, tempo	1192
2	afeganistao, ajudar, acontecer, gente, deixar, triste, precisar, pensar, populacao, coisa	1190
3	brasil, mundo, chegar, eua, cuba, viver, pais, acabar, contra, levar	1098
4	pessoa, atingir, ficar, apos_terremoto, terremoto_magnitude, matar, militar, onu, vida, morte	365

Tabela 52 – Tópicos de agosto de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, presidente, ajudar, terremoto, grande, pessoa, gente, venezuela, deixar, passar	3763
1	haitiano, eua, governo, biden, imigrante, acabar, chegar, enviar, pedir, americano	716
2	brasil, argentina, afeganistao, bolsonaro, inflacao, america_latina, militar, apenas, perder, pobre	80
3	povo, ficar, mundo, onu, pais, politico, cuba, primeiro, pensar, trabalho	20
4	brasileiro, crianca, lembrar, haitiana, receber, eua_deportar, esperar, deportar, apoio, trump	20

Tabela 53 – Tópicos de setembro de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, eua, governo, pessoa, presidente, venezuela, americano, mundo, povo	2498
1	haitiano, revolucao, militar, bolsonaro, enviar, escola, tentar, assassino_presidente, receber, caso	97
2	ficar, grande, matar, fao, primeiro, projeto, conhecer, capital, voltar, apenas	30
3	argentina, pais, gente, mau, inflacao, venezuela, coisa, boliver, america_latina, olhar	24
4	brasileiro, contra, crianca, deportar, sair, mandar, gangues, sequestro, situacao, parte	13

Tabela 54 – Tópicos de outubro de 2021.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, brasileiro, grande, mundo, ficar, povo, lula, negro, pais	1927
1	governo, general, presidente, cuba, pobre, bolsonaro, passar, conseguir, morrer, haitiana	53
2	haitiano, revolucao, pessoa, venezuelano, sair, entrar, seguir, morar, casa, negro	34
3	venezuela, eua, politico, pai_pobre, america_latina, precisar, ganhar, fechar, chegar, acabar	18
4	gente, chegar, heleno, deixar, ajudar, acabar, pensar, acontecer, estar, matar	9

Tabela 55 – Tópicos de novembro de 2021.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	haitiano, brasil, lula, brasileiro, governo, presidente, mundo, cuba, revolucao, acontecer	1538
<b>1</b>	ficar, heleno, matar, general, pais, pobre, militar, deixar, levar, contra	236
<b>2</b>	grande, gente, eua, general_heleno, bolsonaro, genocida, exercito, chegar, pensar, chamar	214
<b>3</b>	venezuela, passar, pai_pobre, coisa, tornar, mandar, america, america_latina, entender, tempo	91
<b>4</b>	pessoa, povo, ajudar, explosao_caminhao, conseguir, lugar, senhor, assassinar, crianca, trabalhar	40

Tabela 56 – Tópicos de dezembro de 2021.

## E APÊNDICE - TÓPICOS EXTRAÍDOS NOS CINCO MESES DO ANO DE 2022

Tópicos	Palavras	Documentos associados
0	haiti, brasil, eua, presidente, ficar, pessoa, mundo, terremoto, governo, independencia	2100
1	haitiano, grande, revolucao, haitiana, acabar, negro, revolucao, demais, america, ouvir	110
2	cuba, pais, brasileiro, argentina, primeiro, gente, venezuela, america_latina, pequeno, populacao	39
3	contra, povo, militar, bolsonaro, pobre, venezuela, selecao, rico, ajudar, brasileiro	22
4	primeiro_ministro, usar, passar, sair, chegar, entrar, coisa, ariel_henry, contar, imigrante	3

Tabela 57 – Tópicos de janeiro de 2022.

Tópicos	Palavras	Documentos associados
0	brasil, eua, cuba, guerra, mundo, pais, venezuela, lembrar, congo, ucrania	638
1	brasileiro, gente, militar, governo, onu, negro, primeiro, pensar, lula, matar	573
2	haitiano, imigrante, africano, acontecer, revolucao, missao, coisa, levar, meter, viver	309
3	ficar, povo, jamaica, ficar_africa, ajudar, passar, mandar, mau, vida, perguntar	272
4	africa, pessoa, comentario, grande, presidente, responder, pessoal, colocar, todo_mundo, politico	194

Tabela 58 – Tópicos de fevereiro de 2022.

Tópicos	Palavras	Documentos associados
0	haiti, brasil, guerra, brasileiro, eua, cuba, ucrania, pessoa, grande, gente	2162
1	haitiano, passar, negro, imigrante, chamar, ucraniano, conseguir, venezuelano, perguntar, trabalho	71
2	mundo, ficar, pais, militar, governo, lula, contra, coisa, invadir, congo	23
3	onu, lembrar, crianca, primeiro, terremoto, verdade, ganhar, proprio, historia, general_heleno	5
4	ajudar, mulher, pobre, lugar, mau, apenas, outro_pais, presidente, parte, filho	2

Tabela 59 – Tópicos de março de 2022.

Tópicos	Palavras	Documentos associados
0	haiti, brasileiro, ficar, pessoa, povo, coisa, passar, venezuela, conseguir, servir	1874
1	haitiano, revolucao, imigrante, chegar, mau, venezuelano, trabalhar, chamar, tomar, olhar	97
2	brasil, eua, grande, guerra, mundo, cuba, pais, contra, deixar, missao	89
3	militar, exercito, onu, governo, lula, general, acabar, presidente, caso, pobre	33
4	gente, negro, matar, heleno, igual, precisar, politico, colocar, acontecer, verdade	11

Tabela 60 – Tópicos de abril de 2022.

<b>Tópicos</b>	<b>Palavras</b>	<b>Documentos associados</b>
<b>0</b>	brasil, venezuela, bolsonaro, cuba, pais, mau, franco, gente, grande, governo	760
<b>1</b>	haitiano, ficar, povo, revolucao, guerra, lembrar, coisa, acabar, acontecer, branco	641
<b>2</b>	onu, brasileiro, pessoa, mandar, missao_paz, paz, usar, missao, america, precisar	276
<b>3</b>	mundo, presidente, crescer, primeiro, pequeno, conseguir, independencia, pensar, deixar, ajudar	186
<b>4</b>	lula, eua, contra, militar, heleno, general, amigo, viver, chegar, golpe	185

Tabela 61 – Tópicos de maio de 2022.