

Fernanda Rocha de Moraes Gonçalves

Feminismos na Internet

Belo Horizonte, Minas Gerais

2021

Fernanda Rocha de Moraes Gonçalves

Feminismos na Internet

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: Mirella M. Moro

Belo Horizonte, Minas Gerais
2021

Agradecimentos

Agradeço primeiramente aos meus pais, que me ajudaram e me incentivaram nas minhas escolhas desde o início e, principalmente, por serem exemplos de profissionais na área de Tecnologia da Informação, nos quais me inspiro todos os dias. À minha mãe por me dar a oportunidade de descobrir a área de gestão e ao meu pai por estar sempre disposto a me ajudar com seu conhecimento técnico. Agradeço muito à minha irmã, meu porto seguro de apoio. Agradeço imensamente ao meu avô, por sempre me distrair quando eu precisava e por me incentivar a ser estudiosa.

Quero agradecer também à minha trajetória dentro da empresa júnior do DCC, a iJunior, que com certeza me mostrou que escolhi o curso certo. E agradeço aos meus colegas de turma, pelo apoio e pela troca de conhecimento. Aos meus professores, por todos os ensinamentos.

Resumo

A análise de dados manualmente é um trabalho exaustivo, que nem sempre recupera todos os insights possíveis dos dados, independente do tema a ser analisado. Dado a quantidade de dados da internet, esse trabalho fica ainda mais desafiador. Nesse contexto, buscou-se analisar os dados dos posts de um blog de cunho feminista com o apoio de algoritmos de análise de texto e identificação de tópicos a fim de gerar visualizações que representem os dados e facilitem uma análise qualitativa e quantitativa de maneira mais automatizada. Assim, foi possível estabelecer relações de frequência entre os temas do site e obter uma visão geral e temporal das publicações. Entretanto, nem todos os algoritmos aplicados geraram resultados relevantes e alguns evidenciaram a dificuldade em analisar textos em português.

Palavras-chave: feminismo, blog, web crawling, web scraping, análise de dados

Abstract

Manually analyzing data is exhausting work, which does not always recover all possible insights from the data, regardless of the topic to be analyzed. Given the amount of data on the internet, this work is even more challenging. In this context, we seek to analyze data from posts of a feminist blog with text analysis algorithms and topic identification in order to generate visualizations that represent the data and facilitate a qualitative and quantitative analysis in a more automated way. Thus, it was possible to establish frequency relationships between the themes of the site and to obtain an overview and time frame of the publications. However, not all applied algorithms generated relevant results and some showed the difficulty in analyzing texts in Portuguese.

Keywords: feminism, blog, web crawling, web scraping, data analysis

Lista de ilustrações

Figura 1 – Categorias do site	14
Figura 2 – Visão temporal da assinatura dos textos pela fundadora do site	15
Figura 3 – Nuvem de palavras - Conteúdo por ano	15
Figura 4 – Participação percentual dos blocos de palavras nas publicações por ano	16
Figura 5 – Bigramas do conteúdo	17
Figura 6 – Co-ocorrência entre palavras do Bloco 6	17

Lista de tabelas

Tabela 1 – Blocos de palavras	12
Tabela 2 – Autoras e número de publicações autorais	16
Tabela 3 – Palavras dos blocos de interesse que não aparecem no conteúdo do site	18

Sumário

	Lista de ilustrações	5
	Lista de tabelas	6
1	INTRODUÇÃO	8
2	CONTEXTUALIZAÇÃO E TRABALHOS RELACIONADOS	9
3	DESENVOLVIMENTO DO TRABALHO	10
4	RESULTADOS E DISCUSSÃO	13
5	CONCLUSÕES E TRABALHOS FUTUROS	19
	REFERÊNCIAS	20

1 Introdução

O feminismo é um movimento de liberação da mulher que em meados do século XIX iniciou uma luta organizada e coletiva para eliminar as desigualdades entre os sexos [Gamba 2008]. Com o avanço do movimento, os feminismos mostram-se como uma “corrente plural de pensamento e ação” que interpretam de diferentes pontos de vista o movimento feminista [Tavares 2012] .

A internet como meio de comunicação é fundamental para os movimentos sociais contemporâneos [Pereira 2011], sendo a principal forma de manifestação de posicionamento e insatisfações. O feminismo como um movimento social está certamente abarcado no contexto da web.

Neste cenário, visto a quantidade de dados nas plataformas digitais, uma análise manual torna-se cada vez mais difícil. Sendo assim, o objetivo geral deste trabalho é permitir análises históricas e sociais dos feminismos presentes na internet de maneira menos manual e mais automatizada, com o apoio de algoritmos computacionais.

Na primeira parte deste trabalho coletou-se dados e gerou-se visualizações e análises sobre site Lady Comics¹. O projeto do Lady’s Comics foi alimentado durante oito anos, entre 2010 e 2018, sob o lema "HQ não é só para o seu namorado", evidenciando o humor e os quadrinhos como instrumentos de manifestação feminista. Por estar dentro do contexto de blogs, uma plataforma não regulada e sem restrições, as autoras tem a liberdade de divulgar conteúdos ricos e de acordo com uma agenda de interesse das mesmas.

Nessa etapa, a proposta é aplicar um fluxo semelhante de coleta e análise de dados para outra iniciativa de mulheres, o Mina de HQ², também em formato de blog e com divulgações de temas parecidos ao site já analisado. O projeto do Mina de HQ, no ar desde 2015, foi criado pela jornalista Gabriela Borges com foco principalmente na divulgação de quadrinhos de mulheres cisgênero, pessoas transexuais e não binárias

Portanto, planeja-se descobrir como o feminismo aparece nessas publicações, quais são os temas mais recorrentes, como é a distribuição temporal dos temas abordados nos sites, além de buscar semelhanças e diferenças entre os dois blogs que foram analisados. Ou seja, busca-se entender as dinâmicas entre os feminismos e a internet.

¹ <http://ladyscomics.com.br/>

² <https://minadehq.com.br/>

2 Contextualização e trabalhos relacionados

Em pesquisa feita por [Bandeli, Hussain e Agarwal 2020], os autores propõem um framework para análise de dados contidos em um blog que utiliza técnicas de processamento de linguagem natural para identificar temas e narrativas presentes. Sendo assim, evidencia-se a importância do uso de algoritmos computacionais na análise da quantidade massiva de dados presente na internet.

No artigo de [Raghupathi, Zhou e Raghupathi 2019], abordagens para a análise de texto não estruturado de blogs são exploradas, utilizando algoritmos de análise de texto em busca de padrões e palavras chave dentro do tema do blog, "Câncer". Já em [Samonte et al. 2017] e [Molin et al. 2019] o foco maior é na análise de sentimentos em publicações de blogs.

Os autores [Elgesem, Feinerer e Steskal 2016], em uma combinação entre métodos automatizadas e manuais, conduzem uma análise de publicações de blogs acerca do caso "Snowden", sendo que o foco dos algoritmos foi identificar os tópicos principais citados nos dados de 15.000 posts.

3 Desenvolvimento do trabalho

Para a obtenção dos dados do site utilizou-se um framework open-source em Python, o Scrapy ¹, que permite fazer o web crawling e o web scraping, de forma a extrair dados estruturados de páginas web. A partir da instalação do framework e da criação do projeto, é necessário configurar as regras de crawling dentro de um Spider. Um Spider é uma classe que define quais páginas web serão percorridas, como ocorrerá a navegação dessas páginas e como serão feitas as extrações dos dados.

Para extrair o conteúdo de interesse, o Scrapy usa uma função de callback para percorrer a árvore de links definidos e os seletores que já estão disponíveis no objeto de resposta do Spider. Nesse projeto, todas as extrações foram feitas usando o seletor XPath (XML Path Language), linguagem que usa um caminho de sintaxe para identificar e navegar por meio dos nós de um documento que utiliza marcadores, como é o caso do HTML. Assim, toda extração de dados é feita a partir do código fonte do site.

O código abaixo exemplifica o uso do XPath, e, nesse caso, seleciona o texto de todas as tags h1 que possuem a classe "entry-title".

```
xpath('//h1[has-class("entry-title")]/text()')
```

Foi gerado, assim, um arquivo CSV a ser analisado, disponível, junto com os Spiders desenvolvidos e demais visualizações geradas, em um repositório do GitHub. ²

Para cada uma das publicação extraiu-se o título da postagem, o conteúdo escrito, a data de publicação, o nome do autor, os comentários e a categoria. Ao configurar o Spider, inclui-se também um valor booleano que indica se a publicação possui ou não imagem, gerando 278 registros extraídos em maio de 2022.

Após a coleta dos dados e criação dos datasets em um Jupyter Notebook, realizou-se a padronização e a transformação dos dados coletados tendo em vista o foco principal de análise dos textos e títulos das publicações. Dessa forma, transformamos as datas das publicações que estavam no formato textual "dia de mês de ano" em datetime, retiramos as pontuações do texto e padronizamos todo o conteúdo em letras minúsculas.

Na geração das nuvens de palavras, duas abordagens foram usadas, sendo a primeira utilizando uma matriz com a frequência das palavras dentro do texto a partir da função FreqDist da biblioteca nltk e a prévia stemmização do texto, e a segunda diretamente a partir do texto. Notou-se pouca ou nenhuma diferença nas nuvens de palavras geradas a partir das duas abordagens com o mesmo input.

¹ <https://docs.scrapy.org/en/latest/>

² <https://github.com/fernandagonc/blog-analysis>

Ambas as estratégias consideram a definição de "stop words", que são palavras a serem ignoradas pelo algoritmo e envolve, principalmente, conjunções e preposições. Ao gerar algumas nuvens de palavras, observou-se que para análise os substantivos seriam mais significativos, logo retiramos também alguns adjetivos e verbos.

Dado que as historiadoras que acompanham o projeto definiram oito blocos de palavras, cada um com uma temática central - feminismo, raça, humor, entre outras, conforme a Tabela 1 - que constituem palavras de interesse dentro do conteúdo do site, buscou-se gerar visualizações de palavras que evidenciassem a relação dessas palavras com elas mesmas. Para isso, utilizamos novamente a função de distribuição de frequência, porém ao montar o gráfico, selecionamos apenas as palavras definidas previamente.

Buscou-se gerar "n-grams" para análise de contexto das principais palavras encontradas nas nuvens de palavras, uma vez que evidencia como as palavras aparecem juntas nas frases. Para isso, novamente retiramos as "stop words" e usou-se o algoritmo de "n-grams" da biblioteca nltk, que mostra os "n-grams" mais frequentes dentro do conjunto de textos disponível.

Por fim, produziu-se uma matriz de co-ocorrência de palavras a partir dos títulos das publicações. Não foi possível gerar grafos de co-ocorrência a partir dos textos das publicações pois o conteúdo era muito grande e o kernel do Jupyter Notebook falhou em todas as tentativas. A visualização dessa matriz foi feita em forma de grafo, em que as arestas representam o número de co-ocorrência entre as duas palavras que estão nos nós. Os grafos foram gerados usando a biblioteca networkx. Entretanto, para que os grafos fossem legíveis, foi preciso filtrar as palavras que seriam analisadas, de forma que escolheu-se mostrar apenas a relação entre as palavras já definidas nos blocos. Definiu-se ainda, um mínimo de peso 10 para que a relação fosse considerada relevante.

Tabela 1 – Blocos de palavras

Bloco 1	feminismo, feminista, feministas, direitos das mulheres, direitos, igualdade, representação, mulher, personagens mulheres, personagens femininas, mulheres, sororidade, amizades, artistas, artista, mulher artista, desigualdade, desigualdade de gênero, discriminação, preconceito, misoginia, machismo, sexismo, sexualização, estereótipo, corpo, patriarcado, patriarcal, visibilidade, invisibilidade, reconhecimento, padrão, padrões, militância, militar, ativismo, frágeis, papéis de gênero, universo feminino, estereotipada, empoderamento, empoderar, inclusão, opressão, exclusão, marchas, marcha, luta, livre, liberdade, movimento feminista, interseccional
Bloco 2	maternidade, mãe, casamento, violência, abuso, estupro, assédio, sexualidade, trabalho doméstico, doméstico, gravidez, aborto, sexo, masturbação, prazer, pornografia, sensual, sensualidade, tesão, nudes, libido, liberdade sexual, política, trabalho, mercado de trabalho, capitalismo, consumo, amor, relacionamentos, afetos, feminilidade, pobreza, privilégio
Bloco 3	raça, racismo, antirracismo, etnia, indígena, índio, negro, negra, lésbica, lesbianismo, transexual, travesti, transgênero, mulher trans, mulheres trans, homem trans, homens trans, identidade de gênero, identidade, masculino, masculina, homens, homem, LGBT, LGBTQIA+, LGBTTFobia, não-binário, heteronormatividade, homoafetividade, deficiência, capacitismo, queer, diversidade, religião, religiosidade, família
Bloco 4	humor, riso, ironia, cômico, comicidade, cômica, engraçado, engraçada, graça, rir, sarcasmo, humorada, humorístico, divertido, divertida, absurdo, horror, terror, medo, caos, pânico, grotesco
Bloco 5	história, memória, passado, internet, web, webcomic, e-comic, redes sociais, instagram, facebook, whatsapp, youtube, zap, site, blog, twitter, mídias sociais, imprensa, livros, editoras, revistas, jornais, resgatar, protagonista, protagonismo, recordar, recuperar, tecnologia, processo criativo, criação, criatividade, cultura, escrita, texto, literatura, literário, linguagem, autobiografia, auto ficção, técnica, Marvel, DC, Disney, MSP, Turma da Mônica, Maurício de Sousa, princesa, heroína, herói, mainstream, prêmio, premiação, evento, Comic-Con, CCXP, HQMIX, nerd, troféu, indicação, concurso, estudo, estudar, formação, curso, teoria, teórico, plágio
Bloco 6	covid-19, covid, coronavírus, vírus, vacina, pandemia, quarentena, doença, morte, isolamento, máscara, distanciamento, solidão, saudade, home office, rotina, saúde mental, crise, autocuidado, autoconhecimento, ansiedade, produtividade, burnout, autoconfiança, autoestima, novo normal, bolsonaro, ele não, bozo, política, político, economia, resistência, discurso de ódio, medo, fascismo, desinformação, fake news, eleições, eleição, esperança, empatia
Bloco 7	educação, ensino, escola, infância, criança, infantil, adolescente, adolescência, professor, professora, pedagogia, didática
Bloco 8	América Latina, estrangeiro, estrangeira, Brasil, brasileira, brasileiro, norte, nordeste, sul, sudeste, centro-oeste, Acre, Alagoas, Amapá, Amazonas, Bahia, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte, Rio Grande do Sul, Rondônia, Roraima, Santa Catarina, São Paulo, Sergipe, Tocantins

4 Resultados e Discussão

Os resultados deste trabalho consistem em visualizações dos dados coletados do site, evidenciadas a seguir. Os dados foram coletados em maio de 2022.

Haviam sido definidos previamente categorias de interesse dentro do site: Artigos, Dicas e resenhas de quadrinhos, Fixo 1 (Banco de quadrinistas), Mina Convida, Quadrinhos para ler Artistas para conhecer e HQs para apoiar. Entretanto, das 12 categorias associadas às publicações, evidenciadas no Gráfico 1, não há correspondência entre as três últimas. Sendo assim, em uma divisão de categorias, será dado destaque às outras seis elencadas.

A Tabela 2 aponta o número de contribuições de cada autora para o site. Nota-se que uma das autoras que mais contribuíram para o Lady's Comics, a Samanta Coan, já fez uma publicação no Mina de HQ. Observa-se ainda que a grande maioria dos textos foi feito pela fundadora do site, assinando às vezes como Mina de HQ e às vezes como a própria Gabriela Borges.

Ao analisarmos as assinaturas dos textos feitos pela fundadora do site, a fim de buscar algum motivo pela diferenciação na autoria dos textos, temos que em 2019 não há nenhuma publicação assinada por "Mina de HQ" enquanto que nos anos seguintes o número de assinaturas como Mina de HQ aumenta significativamente e o número de assinaturas como Gabriela vem diminuindo, conforme o Gráfico 2.

Além disso, a diferença na escolha da assinatura não parece associada a categorias ou temas de publicações específicas. Das 21 publicações assinadas como Gabriela Borges, a partir de 2020, há diferentes tipos de publicações.

Diversas nuvens de palavras foram geradas para análise do conteúdo das publicações, separando as visualizações por ano, por categoria e por bloco de palavras. Na figura 3, observa-se a nuvem de palavras separada por ano, com destaque colorido para palavras que fazem parte de algum dos blocos de interesse, de modo que cada cor representa um dos blocos. As palavras em preto são aquelas que não estão nos blocos de interesse. As cores estão evidenciadas na legenda da figura 4, que mostra a prevalência percentual dos blocos de palavras com o passar dos meses no site.

Essa visualização permite observar como os temas apareceram ao longo do tempo e em que períodos há mais destaque a algum deles. Nota-se que a frequência de da maioria dos blocos é semelhante durante os meses, de forma que os blocos 1 e 2 aparentam ser os mais frequentes.

A partir das nuvens de palavras, foi possível notar um destaque grande da palavra "trabalho". Para entender um pouco mais em que contexto essa palavra vinha associada, levantou-se os mais frequentes bigramas dentro do material de análise, desconsiderando as "stop words". Conforme a Figura 5 gerado, nota-se que as duas palavras mais associadas a "trabalho" são "independente" e "financeiramente".

Ademais, fica evidente que a palavra "história" se mantém relevante nas nuvens de palavras e, a princípio, imaginou-se que seria por acompanhar a palavra "quadrinho". Entretanto, a Figura 5 evidencia que a palavra que acompanha "quadrinhos" frequentemente é a variação no plural, "histórias". Isso sugere uma importância independente da variação da palavra no singular.

Ao gerar nuvens de palavras com apenas as palavras de cada bloco de interesse foi possível perceber que várias das palavras que estavam nos blocos sequer apareciam nas nuvens, ou seja, também não estavam presentes no conteúdo do site. Sendo assim, foi feito um levantamento de quais eram as palavras de interesse que nunca apareceram no conteúdo, evidenciadas na tabela 3.

Por fim, com o objetivo de estabelecer uma relação entre os blocos, foram feitos grafos que relacionam a co-ocorrência das palavras de um bloco A com as palavras de um bloco B ou mesmo com as próprias palavras do bloco A. Uma aresta preta representa uma co-ocorrência maior, enquanto que uma aresta azul representa uma menor co-ocorrência. A co-ocorrência indica o número de vezes em que duas palavras aparecem no mesmo contexto, permitindo a associação entre elas. Na Figura 6, observa-se uma forte relação entre pandemia e quarentena.

Figura 1 – Categorias do site

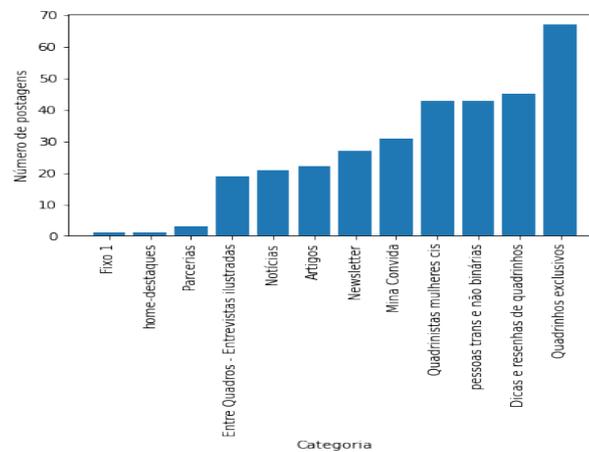


Figura 2 – Visão temporal da assinatura dos textos pela fundadora do site

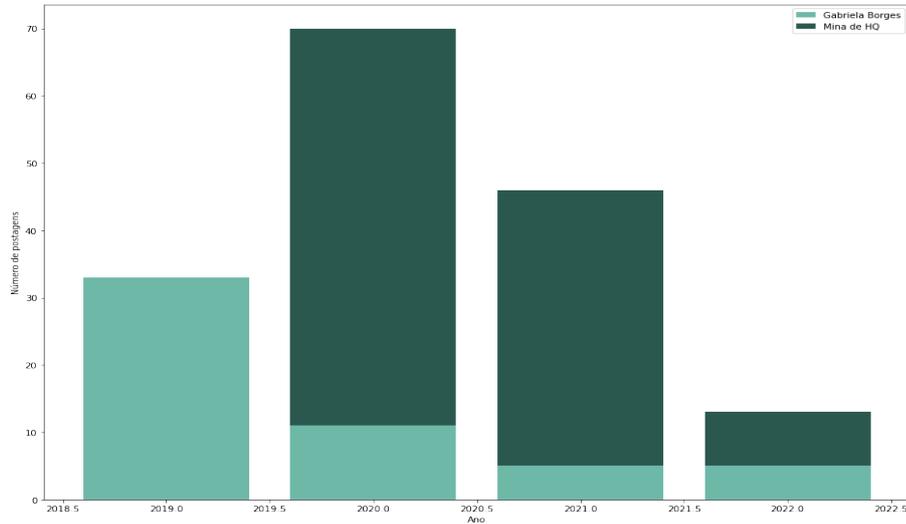


Figura 3 – Nuvem de palavras - Conteúdo por ano



Tabela 2 – Autoras e número de publicações autorais

Autora	Publicações
Mina de HQ	108
Gabriela Borges	54
Gabriella Güllich	18
Anne Ribeiro	17
Sâmela Hidalgo	15
Vitorelo	11
Ellie Irineu	11
Natália Sierpinski	9
Larissa Camargo	5
Dani Marino	5
Maria Júlia Lledó	4
Mayara Lista	3
Kel spinelli	2
Carol Ito	2
Nara Bretas	2
Simone A. Fernandes Anastácio	1
Samanta Coan	1
Redação	1
Ana Fernandes	1
Monique Malcher	1
Maria Clara Villas	1
Lia Bloc	1
Cecília Marins	1
Mari Pelli	1

Figura 4 – Participação percentual dos blocos de palavras nas publicações por ano

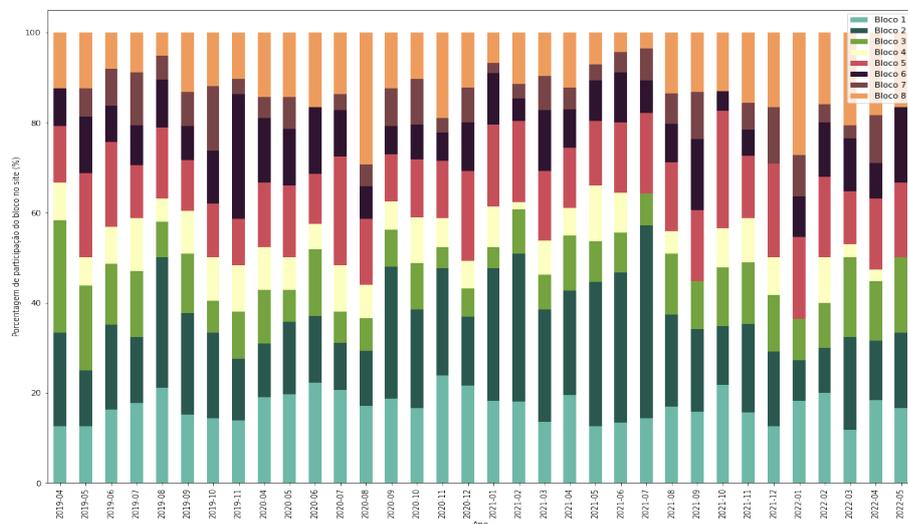


Figura 5 – Bigramas do conteúdo

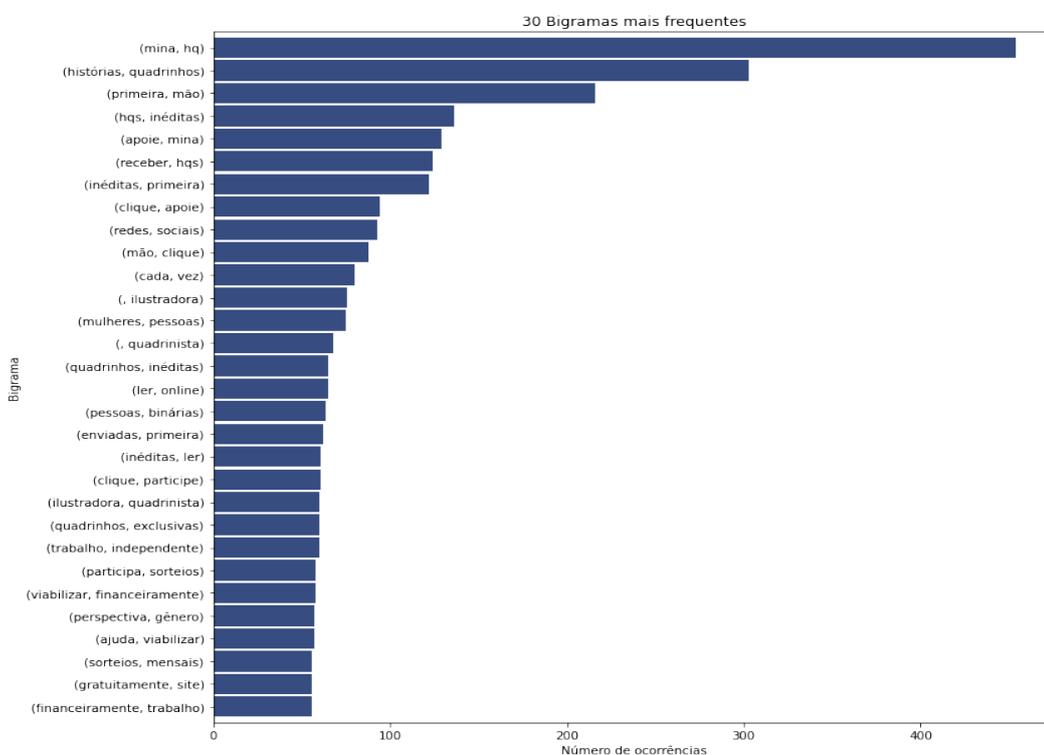


Figura 6 – Co-ocorrência entre palavras do Bloco 6

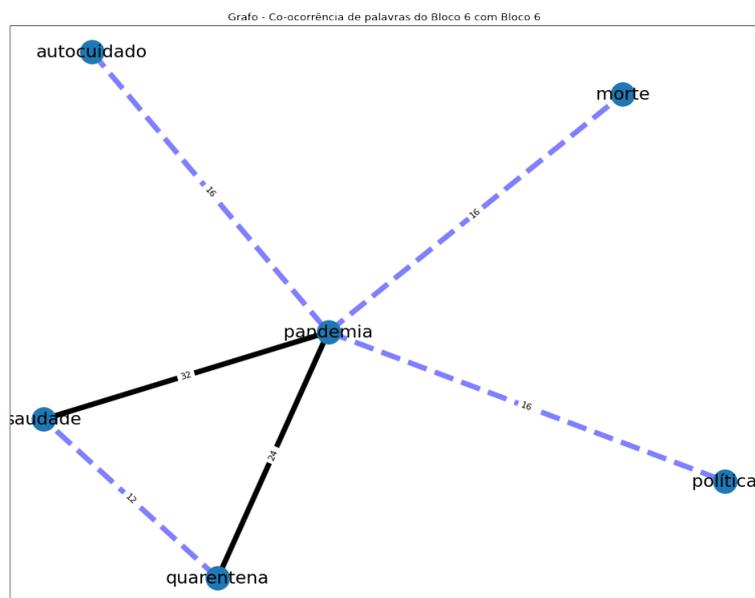


Tabela 3 – Palavras dos blocos de interesse que não aparecem no conteúdo do site

Bloco 1	direitos das mulheres, personagens mulheres, personagens femininas, sororidade, mulher artista, desigualdade de gênero, ativismo, papéis de gênero, universo feminino, empoderar, marchas, marcha, movimento feminista
Bloco 2	trabalho doméstico, sensual, nudes, liberdade sexual, mercado de trabalho, pobreza
Bloco 3	índio, lesbianismo, mulher trans, mulheres trans, homem trans, homens trans, identidade de gênero, LGBT, LGBTQIA+, LGBTTfobia, não-binário, homoafetividade
Bloco 4	comicidade
Bloco 5	e-comic, redes sociais, mídias sociais, recordar, processo criativo, auto ficção, Marvel, DC, Disney, MSP, Turma da Mônica, Maurício de Sousa, Comic-Con, CCXP, HQMIX, teórico, plágio
Bloco 6	covid-19, home office, saúde mental, burnout, autoconfiança, novo normal, ele não, bozo, discurso de ódio, fake news, eleição
Bloco 7	pedagogia
Bloco 8	américa latina, centro-oeste, acre, amapá, distrito federal, espírito santo, mato grosso, mato grosso do sul, minas gerais, paraná, rio de janeiro, rio grande do norte, rio grande do sul, roraima, santa catarina, são paulo, tocantins

5 Conclusões e trabalhos futuros

Durante o desenvolvimento desse trabalho outras visualizações também foram geradas a fim de extrair direcionamentos para as análises, principalmente com diversas nuvens associadas a uma legenda de cores por bloco de interesse e até mesmo nuvens por categoria.

Uma das limitações observadas da análise feita neste trabalho é a restrição de usar como dado apenas o conteúdo textual de um blog que possui muitas imagens. Sendo assim, algumas informações cruciais para o entendimento de questões que foram levantadas podem estar nessas imagens.

Por outro lado, um resultado positivo foi o feedback que a análise trouxe para as historiadoras que propuseram o projeto, uma vez que as visualizações apontam insights que não haviam sido notados somente com uma análise manual. Com uma nova perspectiva de conteúdo, fazer uma nova leitura manual pode ser extremamamente rico.

Das visualizações geradas, foi possível observar a prevalência dos temas no site, uma visão temporal das publicações, além de uma visão geral do conteúdo sem necessidade de lê-lo manualmente. Sendo assim, nota-se a importância dos algoritmos computacionais na extração de insights de uma quantidade notória de dados. Além disso, ao analisar o conjunto de dados sem um filtro das palavras de interesse, as nuvens de palavras acabaram se tornando pouco significativas, ou seja, uma pré-definição dos assuntos e de interesse pelas historiadoras ajudou a direcionar quais seriam as melhores abordagens frente aos dados coletados.

Referências

- [Bairros 1995]BAIRROS, L. Nossos feminismos revisitados. *Revista Estudos Feministas*, v. 3, n. 2, p. 458–458, 1995.
- [Bandeli, Hussain e Agarwal 2020]BANDELI, K. K.; HUSSAIN, M. N.; AGARWAL, N. A framework towards computational narrative analysis on blogs. In: CAMPOS, R. et al. (Ed.). *Proceedings of Text2Story - Third Workshop on Narrative Extraction From Texts co-located with 42nd European Conference on Information Retrieval, Text2Story@ECIR 2020, Lisbon, Portugal, April 14th, 2020 [online only]*. CEUR-WS.org, 2020. (CEUR Workshop Proceedings, v. 2593), p. 63–69. Disponível em: <<http://ceur-ws.org/Vol-2593/paper8.pdf>>.
- [Elgesem, Feinerer e Steskal 2016]ELGESEM, D.; FEINERER, I.; STESKAL, L. Bloggers' responses to the snowden affair: Combining automated and manual methods in the analysis of news blogging. *Comput. Support. Cooperative Work.*, v. 25, n. 2-3, p. 167–191, 2016. Disponível em: <<https://doi.org/10.1007/s10606-016-9251-z>>.
- [Gamba 2008]GAMBA, S. Feminismo: historia y corrientes. *Diccionario de estudios de Género y Feminismos*, v. 3, p. 1–8, 2008.
- [Guzmán 2018]GUZMÁN, M. I. S. Feminismo y tecnología: Software libre y cultura hacker como medio para la apropiación tecnológica. In: PALAU, J. G.; RUSSELL, I. G. (Ed.). *13th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2018, Mexico City, Mexico, June 26-29, 2018, Conference Abstracts*. Red de Humanidades Digitales A. C., 2018. p. 285–286. Disponível em: <<https://dh2018.adho.org/en/feminismo-y-tecnologia-software-libre-y-cultura-hacker-como-medio-para-la-apropiacion-tecnologica/>>.
- [Hartwig e Dearing 1979]HARTWIG, F.; DEARING, B. E. *Exploratory data analysis*. London: Sage Publications, 1979.
- [Molin et al. 2019]MOLIN, G. P. D. et al. Cross-media sentiment analysis in brazilian blogs. In: BEBIS, G. et al. (Ed.). *Advances in Visual Computing - 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7-9, 2019, Proceedings, Part II*. Springer, 2019. (Lecture Notes in Computer Science, v. 11845), p. 492–503. Disponível em: <https://doi.org/10.1007/978-3-030-33723-0_40>.
- [Myers e McGuffee 2015]MYERS, D.; MCGUFFEE, J. W. Choosing scrapy. *Journal of Computing Sciences in Colleges*, Consortium for Computing Sciences in Colleges, v. 31, n. 1, p. 83–89, 2015.

- [Pereira 2011]PEREIRA, M. A. Internet e mobilização política: os movimentos sociais na era digital. *Encontro da compolítica*, v. 4, p. 1–26, 2011.
- [Pimentel et al. 2021]PIMENTEL, J. F. et al. Ciência de dados com reprodutibilidade usando jupyter. In: ANDRADE, A. M.; WAZLAWICK, R. S. (Ed.). *Jornada de Atualização em Informática 2021*. Online: SBC, 2021.
- [Raghupathi, Zhou e Raghupathi 2019]RAGHUPATHI, V.; ZHOU, Y.; RAGHUPATHI, W. Exploring big data analytic approaches to cancer blog text analysis. *Int. J. Heal. Inf. Syst. Informatics*, v. 14, n. 4, p. 1–20, 2019. Disponível em: <<https://doi.org/10.4018/IJHISI.2019100101>>.
- [Randles et al. 2017]RANDLES, B. M. et al. Using the jupyter notebook as a tool for open science: An empirical study. In: *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. Toronto, Ontario, Canada: ACM, 2017. p. 1–2.
- [Samonte et al. 2017]SAMONTE, M. J. C. et al. Emotion detection in blog posts using keyword spotting and semantic analysis. In: BEN-OTHMAN, J. et al. (Ed.). *Proceedings of the 3rd International Conference on Communication and Information Processing, ICCIP 2017, Tokyo, Japan, November 24-26, 2017*. ACM, 2017. p. 6–13. Disponível em: <<https://doi.org/10.1145/3162957.3162963>>.
- [Tavares 2012]TAVARES, M. *Feminismos: Percursos e desafios*. Alfragide, Portugal: Texto Editores, LDA, 2012.
- [Wolff 2019]WOLFF, C. S. Razón y emoción: mujeres militantes en las dictaduras del cono sur. *Women, Militancy and Violence*, v. 33, 2019.