CaçaFake: A System for Monitoring and Analyzing Low Credibility Websites in Brazil

Márcio Silva*†, Julio C. S. Reis‡, João M. M. Couto*, Leandro Araújo*, João Maduro*, Ana P. C. Silva*, Jussara M. Almeida*, Fabrício Benevenuto*
* Department of Computer Science, Universidade Federal de Minas Gerais (UFMG), Brazil
† College of Computer Science, Universidade Federal do Mato Grosso do Sul (UFMS), Brazil
‡ Department of Infomatics, Universidade Federal de Viçosa (UFV), Brazil
marcio@facom.ufms.br,jreis@ufv.br,{joaocouto,leandroaraujo}@dcc.ufmg.br,jpmm2000@ufmg.br
{ana.coutosilva,jussara,fabricio}@dcc.ufmg.br

ABSTRACT

Combating the spread of misinformation is a complex task. In addition, digital platforms (e.g., social networks, instant messaging apps, etc) enhance the dissemination of content produced by low credibility websites. Thus, monitoring and understanding the main characteristics of these websites has become an important task to interrupt the generation chain and spread of misinformation in our society. In this work, we propose a system called CaçaFake in order to speed up the investigation process by supervisory bodies in the fight against misinformation. Our system displays characteristics associated to websites of low credibility which may be useful to regulatory bodies in their decision-making process, creating a rich resource in the fight against misinformation in Brazil.

Keywords: Fake News, Misinformation, System, Low Credibility Websites

1 Introdução

Nos últimos anos, a produção e a rápida e contínua difusão de informação resultaram em uma nova forma da sociedade se relacionar com os fatos, conhecida como *era da pós-verdade*. Segundo a Academia Brasileira de Letras, o termo pós-verdade pode ser definido como *"informação ou asserção que distorce deliberadamente a verdade, ou algo real, caracterizada pelo forte apelo à emoção, e que, tomando como base crenças difundidas, em detrimento de fatos apurados, tende a ser aceita como verdadeira, influenciando a opinião pública e comportamentos sociais"*[1]. Essa era é caracterizada por um estado de intensa produção e disseminação de desinformação em áreas como saúde pública [5, 13], discurso público e processos eleitorais [1, 6, 8, 12].

Nesse contexto, diversas plataformas digitais, como Twitter, Instagram, WhatsApp e Telegram, se tornaram ambientes amplamente explorados para a difusão de campanhas de desinformação. Por exemplo, esforços anteriores mostraram fortes evidências de que as eleições brasileiras de 2018 foram marcadas pelo uso do *WhatsApp* como meio de difusão de informações e notícias falsas [7, 9, 10]. No entanto, uma parcela significativa do conteúdo de desinformação compartilhado nestas plataformas possui origem em outras fontes, em particular, *websites* externos. Ou seja, em alguns casos, essas plataformas, que incluem mídias sociais e aplicativos de mensagem instantânea, são utilizadas apenas como mecanismos para propagação de determinado conteúdo. De forma geral, estes *websites*, reais produtores de conteúdo, podem ser classificados como de *alta credibilidade* ou de *baixa credibilidade* [11]. Aqui, baseado em esforços anteriores [2, 3] consideramos de *baixa credibilidade* um *website* que tenha publicado algum conteúdo classificado como instância de desinformação por alguma agência de verificações de fatos membro da *The International Fact-Checking Network*(IFCN)[2]. Por outro lado, podem ser considerados como *websites* de *alta credibilidade* aqueles membros da Associação Nacional de Jornais (ANJ)[3].

Assim, ao considerarmos a disseminação de conteúdo de desinformação, a criação e manutenção de *websites* para hospedar conteúdos de *baixa credibilidade* passam a ser fundamentais para o sucesso da propagação deste tipo de conteúdo. Muitos destes *websites* mimetizam a estrutura de grandes portais de notícias para passar credibilidade como veículo de comunicação a potenciais leitores. Logo, monitorar e entender as principais características destes *websites* tornou-se uma tarefa extremamente relevante para interromper a cadeia de geração e difusão de desinformação em nossa sociedade.

Diante deste contexto, este trabalho apresenta o CaçaFake, um sistema que realiza o monitoramento e análise de informações publicadas e/ou relacionadas a *websites* de baixa credibilidade. Dentre as principais características coletadas

[1] https://www.academia.org.br/nossa-lingua/nova-palavra/pos-verdade

[2] https://www.poynter.org/ifcn/

[3] https://www.anj.org.br/

In: XXI Workshop de Ferramentas e Aplicações (WFA 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

ISSN 2596-1683

Figura 1. Visão geral da arquitetura do CaçaFake.

destes *websites* e apresentadas no sistema proposto, destacamos registros DNS, endereços IP e registros de nomes de domínio, certificados TLS, e infraestrutura de hospedagem, que foram investigadas em um estudo anterior [3]. De forma geral, acreditamos que o sistema proposto possa ser útil em diversos contextos, como por exemplo, auxiliando na identificação precoce de (características de) um *website* focado na produção e disseminação de desinformação em plataformas digitais. Por fim, é válido ressaltar que não é objetivo deste trabalho fornecer uma lista de *websites* de baixa credibilidade, mas sim, oferecer aos pesquisadores e orgãos compententes, que investigam a dinâmica de propagação de desinformação no Brasil, uma ferramenta para entendimento do fenômeno bem como, eventualmente, apoiar à tomada de ações cabíveis (i.e., investigação).

Este artigo está organizado conforme descrito a seguir. Na Seção 2 apresentamos a arquitetura do sistema proposto. Em seguida, requisitos para execução e uma apresentação do sistema são apresentados nas Seções 3 e 4, respectivamente. Por fim, na Seção 4 concluímos o trabalho e apresentamos direcionamentos para pesquisas futuras.

2 Arquitetura do Sistema

O CaçaFake possui três componentes principais: (*i*) módulo de *websites* de baixa credibilidade, o (*ii*) módulo coletor de *snapshots* do conteúdo disseminado pelos *websites* monitorados e, por fim, um (*iii*) módulo contendo uma aplicação Web que, para cada um dos *websites* monitorados, disponibiliza um conjunto de informações relacionadas. A Figura 1 apresenta uma visão geral da arquitetura do sistema proposto que será detalhada a seguir.

2.1 *Websites* de Baixa Credibilidade

Idealmente, gostaríamos de ter à nossa disposição uma lista validada (e.g., por agências de checagem de fatos, etc) de *websites* brasileiros de baixa credibilidade, responsáveis pela produção e disseminação de desinformação, para permitir, por exemplo, que pesquisadores realizem estudos mais aprofundados para entendimento do problema, e/ou eventualmente, autoridades competentes tomem ações cabíveis dentro deste contexto (e.g., abertura de processo investigativo). Porém, no Brasil, é difícil se obter uma lista de *websites* de baixa credibilidade: agências de checagem evitam apontar fontes que frequentemente compartilham desinformação para evitar, por exemplo, judicialização (ou assédio jurídico). Portanto, infelizmente, tal lista não está disponível, o que nos impulsiona a construir a nossa própria lista. Desta forma, utilizamos a metodologia de construção da lista de *websites* de baixa credibilidade proposta em [2], que baseia-se na interação dos usuários em plataformas digitais. De forma geral, a metodologia se baseia na seguinte premissa: um usuário que compartilha desinformação em algum momento tende a compartilhar conteúdos de desinformação em outros momentos oportunos.

Assim, a partir da base de dados construída e caracterizada em [3], investigamos alternativas para facilitar a interação do usuário final (i.e., especialista) no processo de identificação de *websites* dedicados ao compartilhamento de desinformação no contexto brasileiro. Mais especificamente, nós desenvolvemos o protótipo de uma aplicação Web que permite ao usuário final visualizar e analisar "potenciais" *websites* dedicados à compartilhar desinformação, bem como características relevantes desses *websites* que podem fornecer insumos valiosos para dar suporte ao veredito de um especialista.

2.2 Módulo Extrator de Conteúdo

Uma atividade comum em *websites* de baixa credibilidade é a remoção de conteúdo postado. Este fato pode ocorrer, por exemplo, após indícios de que o *website* sofrerá alguma ação judicial ou em decorrência de má repercussão de um conteúdo de veracidade contestável previamente publicado. Esta prática inviabiliza, por exemplo, possíveis investigações conduzidas por autoridades competentes ou a aplicação de penas previstas em lei.

Diante disso, o sistema proposto possui um mecanismo integrado de *snapshots* de conteúdos produzidos por *websites* de baixa credibilidade. Para cada URL de conteúdo contendo uma notícia ou postagem com pelo menos uma checagem de fatos associada, é iniciado um processo de *download* do conteúdo desta URL nos servidores do CaçaFake. Este conteúdo só pode ser acessado via CaçaFake, não sendo possível sua manipulação pelo usuário final da ferramenta. Este processo de *download* e salvamento do conteúdo nos servidores do sistema é realizado a cada 24 horas. Adicionalmente, o sistema cria um link para o *website* archive.org[4], permitindo uma segunda fonte de auditoria pela autoridade competente, conforme exemplo apresentado na Figura 3.

2.3 Módulo Web

Por fim, o módulo Web do CaçaFake utiliza uma arquitetura de microserviços, sendo dividido em duas partes: *frontend* e *backend*. O *frontend* funciona de maneira desacoplada (i.e., microserviço independente) do *backend* do sistema e utiliza

[4]O portal archive.org realiza um versionamento e indexação de conteúdo na Internet.

o padrão REST API[5] para comunicação e *JSON Webtoken*[6] para segurança e autenticação do usuário.

O *frontend* foi implementado em Javascript, HTML e CSS com orquestramento realizado pela biblioteca de interface ReactJS[7]. Já o backend foi escrito com linguagem de programação Python, utilizando o *framework* FastAPI[8] para prover dados ao *frontend* via API REST. Como nossa aplicação possui um *backend* isolado do *frontend*, é possível desenvolver outras interfaces no futuro (e.g.,*mobile*, *desktop*, etc) que se comuniquem com este *backend* sem a necessidade de modificá-lo.

A priori, o usuário terá acesso a uma lista de *websites* de baixa credibilidade. Porém, nossa ferramenta fornece alguns filtros que podem ser utilizados, onde o usuário pode informar intervalo de datas referente ao registro do domínio, palavras-chave, domínio, datas de expiração dos *websites* e pesquisar por nome do proprietário destes domínios. De forma geral, o sistema apresenta uma lista *websites* de baixa credibilidade, indicando de forma visual se o *website* está *online* ou não, por exemplo. Esta verificação de disponibilidade é feita a cada 30 minutos. A tela inicial do sistema, com uma amostra (anonimizada) de *websites* de baixa credibilidade, é apresentada na Figura 2.

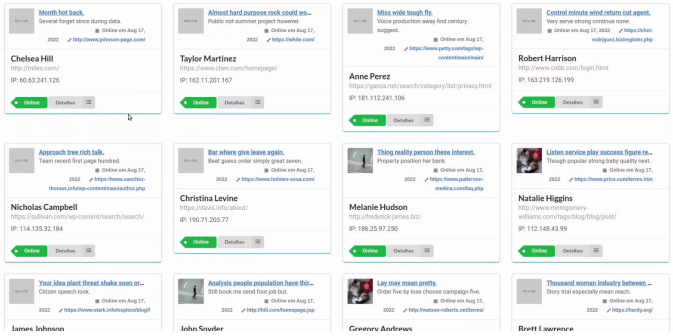

Figura 2. Tela inicial no sistema com lista de *websites* anonimizados.

A Figura 3 mostra a página que permite ao usuário final (especialista) acessar uma cópia local do conteúdo de baixa credibilidade de um *website* específico. Além disso, ele também pode acessar o site archive.org e obter outras versões deste conteúdo. Isso é importante pois é bastante comum a remoção deste tipo de conteúdo (dos *websites* produtores) depois que eles são identificados. Além disso, na Figura 3, nós suprimimos as informações que possam identificar o *website* em questão para evitar judicialização[9], por isso colocamos um *screenshot* do https://br.lipsum.com/ onde seria exibida uma cópia do conteúdo salvo em nossos servidores referente a notícia ou postagem que contém uma checagem que a desminta. Caso o domínio tenha ficado *offline* antes do processo de *download* iniciar, ainda existe a alternativa de clicar no botão "Acessar Agora" para acessar o versionamento de conteúdo feito pelo portal archive.org.

Figura 3. Tela do sistema onde é possível auditar versões da página com conteúdo de baixa credibilidade. A identificação do *website* foi anonimizada.

Adicionalmente, o CaçaFake apresenta um conjunto de 31 atributos agrupados em três categorias: (*i*) atributos de domínio, (*ii*) atributos de certificado e (*iii*) atributos de geolocalização. Estes atributos estão publicamente disponíveis e podem ser abtidos através de protocolos públicos de consulta (e.g., WHOIS[10] [4]), ou através de serviços comerciais de baixo custo (e.g., IPStack[11]):

- **Domínio:** Atributos relacionados com o registro, operação e configuração do nome de domínio, incluindo dados no DNS;
- **Certificado:** Atributos sobre aspectos de segurança do domínio e *website* extraídos a partir de atributos do certificado TLS (ou ausência dele);
- **Geolocalização:** Atributos obtidos a partir da geolocalizaçao do endereço IP hospedando um *website*.

Maiores detalhes relativos ao processo de implementação e extração de cada um dos atributos exibidos no sistema proposto podem ser obtidos em [3]. Finalmente, em uma análise preliminar, após implantação do sistema (ambiente *online*), foi detectado que 12% dos *websites* da nossa lista

[5]Estilo de arquitetura que define um conjunto de restrições a serem usadas para a criação de *Web services*.
[6]https://jwt.io/
[7]https://pt-br.reactjs.org/
[8]https://fastapi.tiangolo.com/

[9]https://revistaoeste.com/brasil/agencia-de-checagem-aos-fatos-e-condenada-por-publicar-fake-news/
[10]Protocol for querying registration information associated with entities on the Internet: https://who.is.
[11]https://ipstack.com/

de *websites* de baixa credibilidade estavam *offline*[12]. Além disso, outro resultado interessante é que, 9, 7% dos *websites* monitorados removeram o conteúdo considerado enganoso. Portanto, nosso sistema permite identificar que ainda existe uma parcela bastante signicativa de conteúdo desinformativo (91, 3%) ainda disponível (*online*) sendo propagada na Web, o que acreditamos que possa fornecer insumos valiosos para processos investigativos por autoridades competentes.

3 Requisitos para Execução

O CaçaFake foi inteiramente desenvolvido para ser implantado dentro de *containers*, especificamente utilizando a plataforma *Docker*[13], tornando-o escalável e multiplataforma, sendo capaz de processar o grande volume de dados gerados pelas plataformas utilizadas para identificação dos *websites* de baixa credibilidade. Para a sua execução, a plataforma necessita de no mínimo 32GB de memória RAM e 500GB de espaço em disco para armazenar o conteúdo coletado das URLs checadas.

4 Apresentação do Sistema

Um *link* para uma captura de tela narrada do sistema em funcionamento está disponível aqui: https://homepages.dcc.ufmg.br/~fabricio/.

5 Conclusão

Descobrir e manter uma lista de *websites* que propagam desinformação em plataformas digitais e na Web é um problema complexo. Neste trabalho, nós apresentamos um sistema chamado CaçaFake que pode ser útil para auxiliar autoridades competentes a monitorar e/ou investigar características associadas a *websites* de baixa credibilidade, focados na produção e disseminação de conteúdo de veracidade contestável. No entanto, a lista de *websites* de baixa credibilidade não pode ser pública, devido à possibilidade de judicialização (assédio jurídico). Assim, exploramos a metodologia proposta em [2] para criação da lista explorada neste trabalho.

O mecanismo de encontrar novos *websites* de baixa credibilidade a partir de uma semente viabiliza a expansão da lista e fornece uma visão global de uma possível rede de compartilhamento de desinformação, trazendo mais segurança e agilidade na tomada de decisão durante eventos de grande repercussão nacional (e.g., eleições). Logo, embora a ferramenta opere considerando uma lista inicial de *websites* de baixa credibilidade fornecida como entrada, ela é robusta para prover o monitoramento e análise de outros que possam surgir. Por fim, é válido ressaltar que este sistema é parte de um projeto desenvolvido em colaboração com o Ministério Público de Minas Gerais (MPMG), com foco na proposição de abordagens que possam ser úteis para a contenção do problema da desinformação no Brasil.

[12] A lista utilizada neste trabalho foi construída em [3], com base na metodogia apresentada em [2] sendo composta por 41 *websites* de baixa credibilidade.

[13] O Docker (https://www.docker.com/) é um *software* que fornece uma camada de abstração e automação para virtualização do sistema operacional no Windows e no Linux. Em outras palavras, o *Docker* é capaz de encapsular um aplicativo e suas dependências em um recipiente virtual que pode ser executado de forma simplificada em qualquer servidor, permitindo flexibilidade e portabilidade para sua execução, em nuvem pública ou privada ou outros ambientes computacionais.

Agradecimentos

Este trabalho foi parcialmente financiado pelo MPMG, projeto Capacidades Analíticas, CNPQ, FAPEMIG e FAPESP.

Referências

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–236.

[2] Leandro Araújo, Luiz Felipe Nery, Isadora Rodrigues, João M. M. Couto, Julio C.S. Reis, Ana Paula Couto Silva, Jussara M. Almeida, and Fabrício Benevenuto. 2022. Identificando Websites de Desinformação no Brasil. In *Proc. of the Brazilian Symposium on Databases (SBBD)*.

[3] João M. M. Couto, Julio C. S. Reis, Ítalo Cunha, Leandro Araújo, and Fabrício Benevenuto. 2022. Caracterizando Websites de Baixa Credibilidade no Brasil. In *Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*. 503–516.

[4] Leslie Daigle. 2004. *WHOIS protocol specification.* Technical Report.

[5] Anneliese Depoux, Sam Martin, Emilie Karafillakis, Raman Preet, Annelies Wilder-Smith, and Heidi Larson. 2020. The pandemic of social media panic travels faster than the COVID-19 outbreak. *Journal of travel medicine* 27, 3 (2020).

[6] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday* 22, 8 (2017).

[7] Philipe Melo, Johnnatan Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabrício Benevenuto. 2019. WhatsApp Monitor: A Fact-Checking System for WhatsApp. In *Proc. of the Int 'l AAAI Conference on Web and Social Media (ICWSM)*. 676–677.

[8] Julio CS Reis, André Correia, Fabrício Murai, Adriano Veloso, and Fabrício Benevenuto. 2019. Supervised learning for fake news detection. *IEEE Intelligent Systems* 34, 2 (2019), 76–81.

[9] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *Proc. of The Web Conference (WWW)*. 818–828.

[10] Gustavo Resende, Johnnatan Messias, Márcio Silva, Jussara Almeida, Marisa Vasconcelos, and Fabrício Benevenuto. 2018. A system for monitoring public political groups in WhatsApp. In *Proc. of the 24th Brazilian Symposium on Multimedia and the Web (WebMedia)*. 387–390.

[11] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature communications* 9, 1 (2018), 1–9.

[12] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (2017), 150–160.

[13] Sander van Der Linden, Jon Roozenbeek, and Josh Compton. 2020. Inoculating against fake news about COVID-19. *Frontiers in psychology* 11 (2020), 2928.