

Identificação de *fake news*: uma abordagem utilizando métodos de busca e *chatbots*

Yara de Lima Araujo¹, Anderson Cordeiro Chares², Jonice de Oliveira Sampaio³

^{1,2,3}Instituto de Informática
Universidade Federal do Rio de Janeiro (UFRJ)

araujo.yara93@gmail.com, andersoncordeironf@gmail.com,
jonice@dcc.ufrj.br

Abstract. *Given the growth of social media, dissemination of information occurs in a fast and scalable way. This dynamic makes evaluation of informations' veracity a task that consumes time and breaks the social interaction flow. Therefore, rumors or fake news also spread faster on internet. In this paper, we propose a chatbot from Facebook Messenger that retrieves messages' keywords for possible rumors in a dataset, which contains data from Brazilian websites that gather news and classify them into true or fake. To do so, it utilizes some information retrieval techniques.*

Resumo. *Diante do grande crescimento das mídias sociais, a disseminação de informações ocorre de maneira mais rápida e escalável. Esse dinamismo transforma a avaliação da veracidade de uma informação em tarefa que demanda tempo e quebra do fluxo contínuo da interação social. Isto faz com que rumores também se espalhem com maior velocidade na rede. Nesse trabalho, propõe-se a utilização de um chatbot para o Facebook Messenger que, através de técnicas de busca e recuperação da informação, pesquisa em um dataset por palavras-chave de possíveis rumores que recebe de usuários, respondendo com links que possam auxiliar na validação da informação. Esse dataset é composto por conteúdo de sites brasileiros que reúnem notícias classificadas como falsas ou verdadeiras.*

1. Introdução

Nos últimos anos, o exponencial crescimento das mídias sociais intensificou a interação humana na internet permitindo que diversas informações fossem disseminadas rapidamente. Apesar da notória contribuição social trazida pelas tecnologias e ferramentas sociais, a dificuldade em lidar com o excesso de informação contribuiu para o também crescente surgimento de conteúdo não verificado (rumores) e muitas vezes falso (boatos).

A velocidade na comunicação impede que uma avaliação possa ser feita no conteúdo que está sendo trafegado e, além disso, os embates sociais resultantes de divergências sócio-políticas impulsionam a prática da produção de conteúdo duvidoso que sirva de alicerce para críticas ou fomite discussões na rede. Identificar esses boatos se apresenta como um desafio; estabelecer a confiabilidade de informações online é um desafio assustador mas crítico [Conroy, Rubin, & Chen 2015]. De acordo com Vosoughi & Roy [2017], um rumor pode ser definido como uma afirmação não-

verificada que começa em uma fonte (ou mais) e começa a se espalhar ao longo do tempo para diversos nós na rede.

Rumores podem ser classificados em alguns tipos. De acordo com Rubin, Chen & Conroy [2015], três tipos são identificados. Primeiramente, as notícias que provém de jornais sensacionalistas, que fabricam notícias sobre escândalos, pessoas famosas, crimes, com o objetivo de obter audiência; muitas notícias são fabricadas, falsificadas ou exageradas, o que leva a muitos rumores. O segundo tipo são definidos como *hoaxes* (embustes/enganos), rumores criados nas mídias sociais com a finalidade de enganar as pessoas como notícias inverídicas. Estes rumores podem ser validados por engano por meios de notícia tradicionais. Diferente de uma simples brincadeira, *hoaxes* podem causar danos reais a alguém.

Este trabalho está organizado da seguinte forma: a seção 2 resume os trabalhos relacionados, a seção 3 demonstra a aplicação proposta e todas as etapas de seu desenvolvimento, a seção 4 apresenta os resultados preliminares e a seção 5 apresenta a conclusão.

2. Trabalhos Relacionados

Atualmente, diversas técnicas têm emergido no intuito de detectar rumores na internet. O trabalho de Castillo *et al* [2011] utiliza técnicas de aprendizado supervisionado de máquina para classificar um tópico no Twitter como notícia ou conversa pessoal e dentro da classe notícia classificá-la como crédula ou não crédula. A classificação é feita por humanos e um algoritmo de aprendizado extrai os padrões que tweets dessas classes possuem. O trabalho conclui que tópicos de notícias costumam possuir links e uma árvore de propagação maior.

Já Buntain [2015] avalia a crescente de palavras-chave em épocas de crise, relacionando a localização de um determinado evento com o surgimento de mensagens a seu respeito e, utilizando aprendizagem de máquina, analisou a técnica proposta em mídias sociais durante eventos esportivos e adaptou-os a situações de risco como a ocorrência de terremotos. Seus experimentos demonstraram que, dados estes cenários, identificar palavras importantes e verificar suas origens pode auxiliar na avaliação da credibilidade de informação.

No trabalho de Shou *et al.* [2016], os autores desenvolveram uma plataforma chamada *Hoaxy* para coletar notícias de diversas mídias sociais, assim como sites de notícias através de *crawlers* e APIs. Após a coleta, a plataforma somente rastreia atualizações através de *RSS*. Todas as notícias são armazenadas em um banco de dados. Como trabalhos futuros, os autores indicam o desenvolvimento de uma interface web interativa para análise das notícias.

Apesar da variedade de técnicas, ferramentas e plataformas utilizadas para a detecção de rumores em outros países, no Brasil ainda há uma escassez de fontes de dados públicos que possam servir de consulta na tarefa de verificação de informação. Além disso, o controle automático de conteúdo quanto à sua veracidade permanece como uma tarefa difícil às máquinas por requerer níveis de abstração e criatividade inerentes a seres humanos. Sabendo disso, soluções que envolvam o desenvolvimento colaborativo de conhecimento como *crowdsourcing* e o acesso à fontes de dados

abertas, podem servir de alicerce para o desenvolvimento de soluções frente ao desafio da identificação de rumores. Por apresentar esse perfil, acredita-se que a aplicação desenvolvida neste trabalho possa contribuir efetivamente para o problema de identificação de rumores.

3. Aplicação proposta

Nesta seção são apresentados os componentes da aplicação proposta e descritos os processos da criação do dataset de rumores, da API de utilização do dataset e do chatbot desenvolvido para validar a ferramenta de busca.

A Figura 1 apresenta a estrutura da aplicação como um todo. As interações ocorrem com o usuário através do *FakeChatBot*, são recebidas pela API, transformadas em buscas diretas ao dataset de notícias (*busca.py*); após isso resultados são retornados ao usuário. É possível identificar os módulos que compõem todo o processo, como os *crawlers* e a camada na qual o modelo vetorial é desenvolvido, detalhados na seção 3.4.

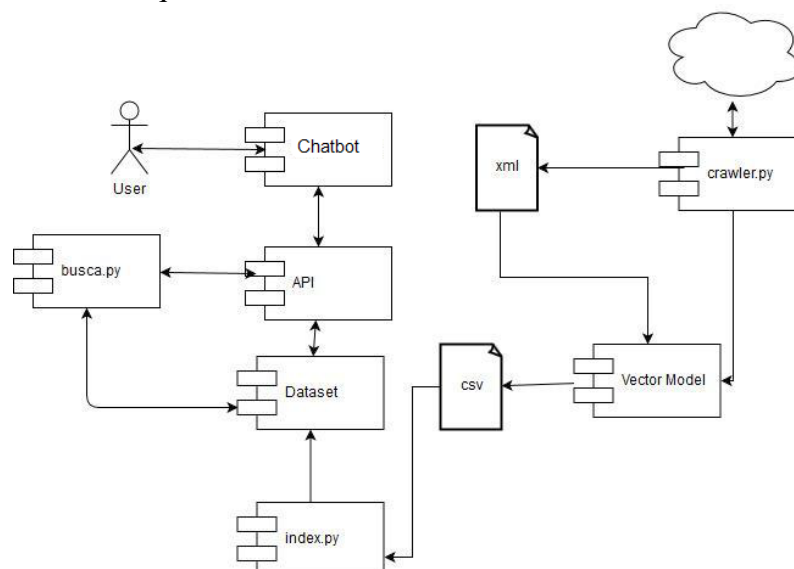


Figura 1. Estrutura da aplicação

3.1 Dataset de Rumores

Com o objetivo de reunir em um *dataset* rumores já conhecidos e assim disponibilizar tal fonte de dados para consultas, foram criados *crawlers* que percorrem sites brasileiros conhecidos por trabalhar com verificação de fatos, fazendo o download do conteúdo e armazenando no *dataset* de maneira estruturada para facilitar a compreensão do conteúdo da notícia e do resultado da avaliação realizada pelo site (falso ou verdadeiro). Foram utilizados os sites “boatos.org”¹, “e-farsas”² e “é ou não é”³. Para realizar esse processo, *crawlers* desenvolvidos em Python extraem os dados dos sites utilizando a biblioteca *Beautiful Soup*, e os escrevem em um arquivo .xml.

O conteúdo dos arquivos .xml é então submetido a técnicas de normalização de texto para que possam ser removidos termos irrelevantes ao mecanismo de busca, como artigos e pronomes *-stop words*. Notícias duplicadas também foram removidas. Após essa etapa, outro *script* é acionado para a criação de um arquivo .csv estruturado com campos que seguem o padrão do .xml como “*summary*” com uma descrição da notícia, e

“*check*”, campo booleano que indica se a notícia foi identificada como verdadeira ou falsa. Após a organização e definição da estrutura, os conteúdo dos arquivos é armazenado no dataset *online* e disponibilizado ao método de busca que está descrito na seção 3.3.

3.2 API

Para facilitar a utilização do dataset, e incentivar o desenvolvimento de aplicações que pudessem utilizá-lo como fonte de dados, foi desenvolvida uma API (*Application Programming Interface*), cujas funções permitem a interação com os registros do dataset.

Através da API é possível listar todos os registros armazenados pelos *crawlers*, utilizar filtros para identificar o que foi classificado como boato ou verdadeiro nos sites especializados e realizar consultas utilizando de um processo bem definido de Busca e Recuperação de Informação. Disponível para acesso em www.fakepedia.org, a API possui uma breve documentação sobre suas funcionalidades e a descrição das chamadas para interação com o *dataset*.

3.3. Métodos de busca e recuperação da informação

Uma das funcionalidades da API a ser utilizada é a busca por um termo ou expressão utilizando o Modelo Vetorial de Busca e Recuperação de Informação. No Modelo Vetorial, um documento é representado por um conjunto de termos indexados e associados a um valor normalizado que indica o seu grau de relevância para o documento. Optou-se pelo Modelo Vetorial por ser bastante difundido e se mostrar uma solução eficaz para o problema de recuperação de texto, como em Senin & Malinchik [2013].

Neste processo, o grau de similaridade (relevância) é calculado através de pesos atribuídos aos termos da consulta e do documento em si. Para o cálculo destes pesos, implementou-se um script que cria uma lista invertida a partir dos arquivos .xml gerados pelos *crawlers*, contendo cada termo e o conjunto de notícias correspondentes. As listas auxiliam no cálculo do TFxIDF dos termos, sendo TF (“*term frequency*”) o número de vezes que um termo aparece em um documento e IDF (“*inverse document frequency*”) a frequência com que um termo ocorre em todo conjunto de documentos. Maiores detalhes sobre TDF-IDF podem ser encontrados em Ramos [2013].

3.4. FakeChatBot

O uso de *chatbots* permite um alcance instantâneo a uma grande quantidade de pessoas [Schlicht 2016]. Dada a popularização e alcance desse tipo de abordagem, o desenvolvimento de um *chatbot* como interface entre o usuário e a API mostrou-se como uma excelente alternativa para a validação do uso da ferramenta de identificação de rumores.

Optou-se por utilizar o Facebook Messenger como plataforma para o desenvolvimento do chatbot devido a sua popularidade, pois é utilizado por mais de um bilhão de pessoas [Constine 2016].

O chatbot, chamado de *FakeCheckBot*, foi desenvolvido em Python, utilizando algumas bibliotecas como *Flask* (*desenvolvimento web*), *requests*, e *unicorn* – espécie

de servidor Python para web. Em seguida, foi criada uma página no *Facebook* de mesmo nome para vincular o chatbot a um perfil da plataforma.

4. Resultados

Após a definição das ferramentas e plataformas a serem utilizadas, bem como a criação da estrutura necessária para receber requisições através do *Facebook Messenger*, testes foram realizados a fim de verificar a qualidade do retorno obtido através das consultas realizadas no *dataset*.

Apesar de carecer de experimentos para validação, resultados preliminares demonstram que o *FakeChatBot* é capaz de identificar trechos de boatos reconhecidos e retornar links com conteúdo já verificado acerca do assunto. A Figura 2 representa o momento em que uma pessoa envia uma pergunta sobre um boato ao *chatbot*, a resposta enviada pelo chatbot após todo o processo de busca e recuperação de informação e métodos aqui descritos.

Notou-se que algumas variações e textos mais longos demandam tempo maior de processamento e, por vezes, não retornam o conteúdo exato por conta das limitações do plano gratuito do servidor utilizado. Também foi notado que a busca de termos muito curtos, contendo uma ou duas palavras, retorna resultados pouco precisos.

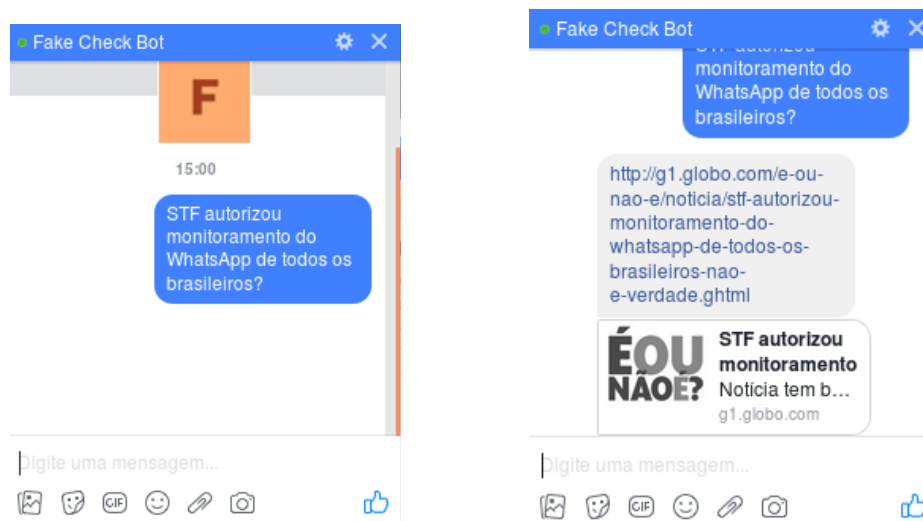


Figura 2. Mensagem enviada pelo usuário e respondida pelo chatbot

5. Conclusão

Diante do crescente uso da internet e das mídias sociais, um grande volume de dados é criado e espalhado rapidamente pela rede. Por consequência, muitos rumores também são disseminados. Entretanto, percebe-se que há dificuldade de avaliação da veracidade de informações por parte de usuários, que não desejam parar suas interações para verificar, por exemplo, se uma notícia é verdadeira.

Neste sentido, muitas abordagens surgiram no intuito de identificar rumores na rede. Neste trabalho, foi proposta uma solução que permite a busca de rumores da internet a partir de um chatbot. Para tal, foram recuperadas informações de sites brasileiros especializados em verificação da informação, através de *crawlers* e reunidos em um único dataset. Para facilitar o acesso, uma API foi desenvolvida e disponibilizada online.

Utilizando Python, foi desenvolvido um chatbot para o *Facebook Messenger* que recebe um texto do usuário e verifica, através da similaridade dos termos, se existe algum boato correspondente no dataset.

Por fim, através desta abordagem foi possível prover uma solução funcional que permite que usuários busquem por rumores de fontes verificadas e descubram se determinada notícia é um rumor ou não. Além disto, propõe-se um *dataset* unificado de bases em português, o que dificilmente está disponível de acordo com um levantamento realizado.

Como trabalhos futuros, pretende-se melhorar o processo de busca através da adoção de técnicas de paráfrase capazes de torná-lo mais inteligente. A partir disto, será possível inferir variações nos termos buscados, que podem ser um mesmo rumor escrito de maneira distinta.

Referências

- Buntain, C. (2015) “Discovering Credible Events in Near Real Time from Social Media Streams”, *WWW 2015 Companion*, May 18–22, 2015, Florence, Italy
- Conroy, N., Victoria L. and Chen, Y. (2015) “Automatic Deception Detection: Methods for Finding Fake News”, *ASIST 2015*, November 610, 2015, St. Louis, MO, USA.
- Ferneda, E. Introdução aos Modelos Computacionais de Recuperação de Informação. [S.l.]: Editora Ciência Moderna, 2012
- Ramos, J. (2003). Using tf-idf to determine word relevance in document queries. *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).
- Rubin, V. L., Chen, Y., & Conroy, N. J. (2015). Deception Detection for News: Three Types of Fake News. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4.
- Senin, P., & Malinchik, S. (2013). Sax-vsm: Interpretable time series classification using sax and vector space model. *IEEE 13th International Conference on Data Mining* (pp. 1175-1180). IEEE.
- Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A Platform for Tracking Online Misinformation. <https://doi.org/10.1145/2872518.2890098>
- Vosoughi, S., & Roy, D. (2017). Rumor Gauge : Predicting the Veracity of Rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*. In Press. 11(4), 1–38. <https://doi.org/10.1145/3070644>