FarolKG: Um Grafo de Conhecimento sobre Desinformação em Mensagens do WhatsApp

Ronney Nery Silva[1], **José Maria Monteiro**[1], **Wellington Franco**[1]

[1]Universidade Federal do Ceará
Fortaleza – CE – Brasil

ronneyns@alu.ufc.br, monteiro@dc.ufc.br, wellington@crateus.ufc.br

Nível: Mestrado
Ingresso: Fevereiro de 2020
Previsão de Término: Fevereiro de 2022
Programa: Programa de Mestrado e Doutorado em Ciência da Computação (MDCC)
Etapas concluídas: Pré-Proposta defendida em Ago/2021 e disciplinas concluídas até Fev/2022.
Defesa da Proposta: Março de 2022 (estimativa)

Abstract. *The popularization of social networks created an environment conducive to the mass spread of misinformation. Currently, the main tool used to spread disinformation is WhatsApp public groups. In this work, we propose the FarolKG, a knowledge graph about the misinformation that is propagated through WhatsApp public groups. The use of FarolKG will facilitate the execution of structured queries, both the message texts and their metadata. FarolKG will be generated through a semi-automatic pipeline, which periodically collects data from public WhatsApp groups, popular fact-checking sites and DBPedia. FarolKG will allow advanced exploration and discovery of information, for example, through queries in the SPARQL language. Additionally, FarolKG will provide a web interface that will enable the construction of queries in a visual way.*

Resumo. *A popularização das redes sociais gerou um ambiente propício para o espalhamento em massa de desinformação. Atualmente, a principal ferramenta utilizada para propagação de desinformação são os grupos públicos do WhatsApp. Neste trabalho, propomos o FarolKG, um grafo de conhecimento (KG) sobre as desinformações que são propagadas por meio dos grupos públicos do WhatsApp. A utilização do FarolKG irá facilitar a execução de consultas estruturadas envolvendo tanto os textos das mensagens quanto os seus metadados. O FarolKG será gerado por meio de um pipeline semiautomático, que coleta dados periodicamente de grupos públicos do WhatsApp, de sites populares de verificação de fatos e da DBPedia. O FarolKG permitirá a exploração avançada e a descoberta de informações, por exemplo, por meio de consultas na linguagem SPARQL. Adicionalmente, o FarolKG irá disponibilizar uma interface Web que possibilitará a construção de consultas de forma visual.*

1. Introdução

A ascensão de redes sociais digitais, onde os conteúdos podem ser produzidos e publicados pelos próprios usuários de forma distribuída, gerou um ambiente propício para o espalhamento em massa de desinformação, o que pode ocasionar problemas em diferentes áreas, tais como saúde e segurança pública, além de colocar em risco a própria democracia. A desinformação pode ser entendida como o processo de produção intencional de um ambiente comunicacional pautado por informações falsas, equivocadas ou descontextualizadas para provocar desordem comunicacional. A desinformação atualiza padrões de manipulação da informação já verificados na imprensa, como a ocultação e a fragmentação de fatos. Porém, possui um impacto ainda maior, tendo em vista que as tecnologias atuais permitem rápida difusão de conteúdos, maior volume, automatização de processos e opacidade em seus fluxos. Além disso, as redes sociais estão cada vez mais inseridas na sociedade e se tornando espaço prioritário para a obtenção de informações.

Vale destacar que, atualmente, no Brasil e em alguns países em desenvolvimento, a principal ferramenta utilizada para propagação de desinformação é o *WhatsApp*. Em fevereiro de 2020, a pesquisa *Panorama Mobile Time/Opinion Box* sobre mensageria móvel no Brasil revelou que o mensageiro está instalado em 99% dos *smartphones* brasileiros [1]. Entre os usuários do aplicativo, 98% disseram que o acessam todos os dias ou quase todos os dias. Em dezembro de 2019, levantamento realizado pelo Congresso Nacional junto a 2.400 pessoas concluiu que 79% das pessoas utilizam o *WhatsApp* como principal fonte de informação [2]. Pesquisa da Fundação Oswaldo Cruz (Fiocruz) mostrou que 73,7% das notícias falsas sobre o novo coronavírus circularam pelo *WhatsApp* [3]. Outros 10,5% foram publicadas no *Instagram* e 15,8% no *Facebook*.

Nesse contexto, a detecção automática de desinformação (*MID - Misinformation Detection*) em mensagens do *WhatsApp* torna-se um desafio crucial. Todavia, devido à natureza privada das mesmas, não há uma grande variedade de métodos desenvolvidos especificamente para este aplicativo. A maioria dos métodos existentes baseia-se na combinação de algoritmos de classificação e técnicas de extração de atributos baseadas em Processamento de Linguagem Natural [de Sá et al. 2021, Cabral et al. 2021, Martins et al. 2021].

Por outro lado, Grafos de Conhecimento ou (*Knowledge Graphs - KG*) estão sendo utilizados como um mecanismo para integrar semanticamente um grande número de fontes de dados heterogêneas em um espaço de dados abrangente. Assim, um KG pode ser construído com a finalidade de integrar dados, fornecer uma visão semanticamente conectada dos dados e prover um acesso integrado por meio do uso de aplicações semânticas [Gómez-Pérez et al. 2017]. Um grafo de conhecimento pode ser definido como um conjunto inter-relacionado de entidades e seus atributos. As entidades são utilizadas para representar objetos do mundo real ou conceitos abstratos, sendo representadas pelos nós do grafo. As conexões entre nós (arestas do grafo) significam relações entre essas entidades e são representadas na forma de triplas (sujeito, relação, objeto).

---

[1] Disponível em: https://blog.opinionbox.com/mensageria-no-brasil-sexta-edicao/

[2] Disponível em: www2.camara.leg.br/a-camara/estruturaadm/ouvidoria/dados/pesquisa-nov-2019-relatorio-completo

[3] Disponível em: https://portal.fiocruz.br/noticia/pesquisa-revela-dados-sobre-fake-news-relacionadas-covid-19

Até onde sabemos, nenhum grafo de conhecimento razoavelmente grande e atualizado de informações estruturadas sobre as desinformações que circulam no *WhatsApp* foi disponibilizado de forma pública. Neste trabalho, buscamos preencher essa lacuna introduzindo o FarolKG, um grafo de conhecimento (KG) sobre as desinformações que são propagadas por meio dos grupos públicos do *WhatsApp*. A utilização do FarolKG irá facilitar a execução de consultas estruturadas envolvendo tanto os textos das mensagens quanto os seus metadados (dia da publicação, hora, estado ou país do usuário, etc). A construção do FarolKG seguirá as recomendações e melhores práticas do W3C. O FarolKG será gerado por meio de um pipeline semiautomático, que coleta dados periodicamente de grupos públicos do *WhatsApp*, de sites populares de verificação de fatos e da DBPedia. O FarolKG permitirá a exploração avançada e a descoberta de informações, por exemplo, por meio de consultas como "encontre todas as mensagens que contêm desinformação e mencionam o STF publicadas em agosto de 2021" ou "encontre os 5 principais políticos mencionados em mensagens que contêm desinformação". Além disso, o FarolKG irá permitir a exploração de dados de várias fontes por meio de consultas SPARQL federadas. Adicionalmente, o FarolKG irá disponibilizar uma interface Web que possibilitará que usuários de fora da comunidade de ciência da computação construam suas consultas e recuperem informações.

O restante deste artigo está organizado da seguinte forma: A Seção 2 destaca os trabalhos relacionados. Na Seção 3, um esboço do *framework* proposto é apresentado. A metodologia utilizada é discutida na Seção 4. Por fim, a Seção 5 conclui este trabalho e apresenta direções para pesquisas futuras.

2. Trabalhos Relacionados

O Farol Digital [de Sá et al. 2021], é uma plataforma cujo objetivo é descobrir grupos públicos do aplicativo *WhatsApp* no Brasil e deles extrair, tratar, anonimizar e analisar as mensagens de texto que circulam nesses grupos. Adicionalmente, a plataforma possibilita a detecção automática de desinformação. Por meio desta plataforma foram construídos três *datasets* de mensagens relacionadas a temas relevantes da atualidade: as eleições de 2018, pandemia de Covid-19 e a vacina para o vírus Sars-CoV-2.

Em [Tchechmedjiev et al. 2019], os autores apresentam uma ferramenta denominada **ClaimsKG**, um grafo de conhecimento construído por um processo semi-automatizado de extração de afirmações e suas revisões em *websites* de *fact-checking* com inquestionável reputação, permitindo exploração e descoberta de informações relevantes por meio de consultas estruturadas sobre a veracidade dessas informações e sobre outros tipos de metadados. Este grafo de conhecimento pode ser acessado por meio de um *triplestore* **Virtuoso** [OpenLink Software 2020] permitindo a execução de consultas escritas na linguagem **SPARQL** [W3C 2013]. Para facilitar seu uso por um público diverso e não especialista, a ferramenta apresenta uma *interface* que permite a execução de consultas e a apresentação dos resultados (Figura 1).

Em [Zhou and Zhang 2004], os autores sugerem uma ontologia para a representação da desinformação, baseando-se em uma análise de suas características. Em [Zhou and Zhang 2007], os autores estendem o trabalho anterior adicionando a construção de uma biblioteca para o aprimoramento do conhecimento a respeito desse tema, além de facilitar o compartilhamento, reuso e gerenciamento dos dados.

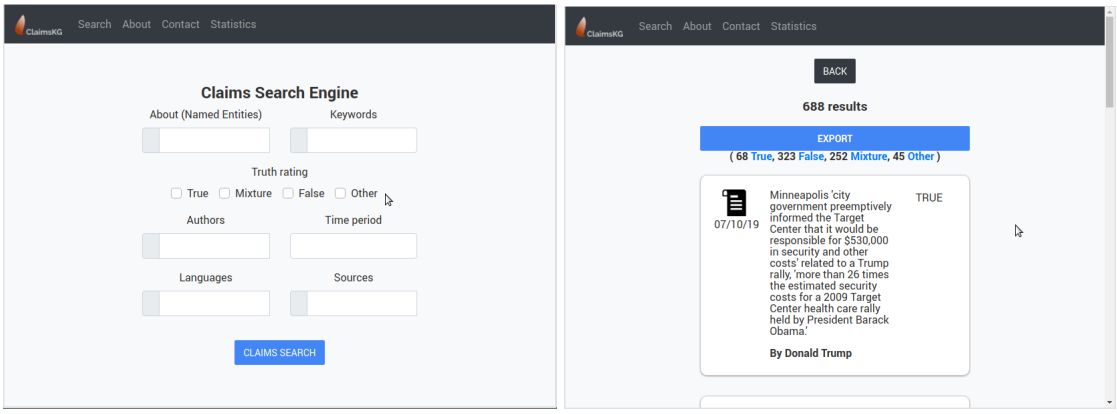

Figura 1. ClaimsKG: *Interface* do Usuário
Fonte: https://data.gesis.org/claimskg/explorer/home

3. O FarolKG

3.1. Visão Geral

Neste trabalho, propomos o FarolKG, um grafo de conhecimento (KG) sobre as desinformações que são propagadas por meio dos grupos públicos do *WhatsApp*. A construção do FarolKG seguirá as recomendações e melhores práticas do W3C. O FarolKG será gerado por meio de um pipeline semiautomático, que coleta dados periodicamente de grupos públicos do *WhatsApp*, de sites populares de verificação de fatos e da DBPedia. A utilização do FarolKG irá facilitar a execução de consultas estruturadas envolvendo tanto os textos das mensagens quanto os seus metadados (dia da publicação, hora, estado ou país do usuário, etc).

Assim, o FarolKG permitirá a exploração avançada e a descoberta de informações, por exemplo, por meio de consultas como "encontre todas as mensagens que contêm desinformação e mencionam o STF publicadas em agosto de 2021" ou "encontre os 5 principais políticos mencionados em mensagens que contêm desinformação". Além disso, o FarolKG irá permitir a exploração de dados de várias fontes por meio de consultas SPARQL federadas. Adicionalmente, o FarolKG irá disponibilizar uma interface da Web que possibilitará que usuários de fora da comunidade de ciência da computação construam suas consultas, de forma visual, e recuperem informações.

3.2. Implementação

A etapa em andamento corresponde ao design das classes e propriedades da ontologia, que está sendo realizada por meio do editor de ontologias **Protégé** [Stanford University 2021], na mais recente versão disponível (5.5).

Na figura 2, apresentamos um esboço da arquitetura a ser desenvolvida para o funcionamento do projeto, com as seguintes camadas:

- Camada de Dados (inicialmente apenas relacional);
- Camada Semântica;
- Camada de Aplicações.

A camada de Dados representa a origem dos dados, no caso em modelo relacional, que irão para o grafo de conhecimento. O banco de dados do Farol Digital será a origem das informações a serem inseridas ou atualizadas periodicamente no FarolKG.

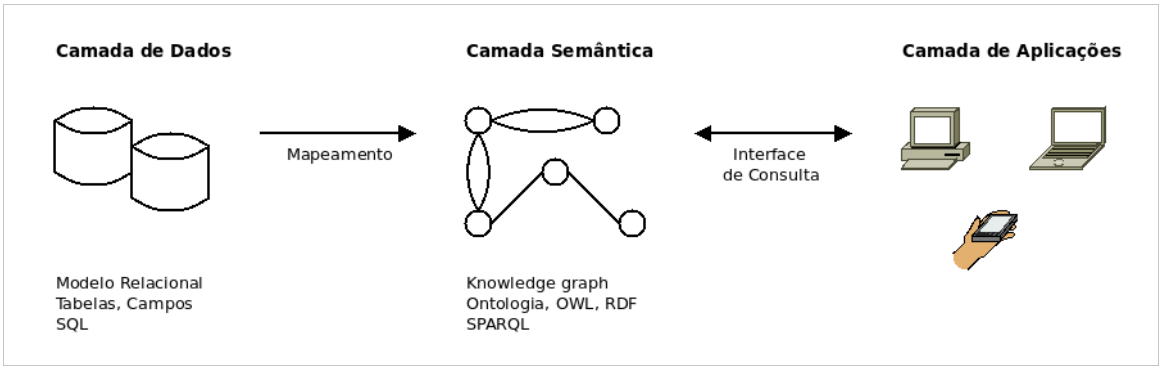

Figura 2. Arquitetura do FarolKG

Entre esta camada e a seguinte teremos um processo de mapeamento dos dados relacionais em classes da ontologia e de identificação de entidades previamente definidas que possam ocorrer nos textos das mensagens, por meio de NLP.

Na camada Semântica, temos a representação na forma de grafo de conhecimento dos dados extraídos usando conceitos de Web Semântica como ontologia, OWL e RDF e a linguagem SPARQL para a realização de consultas.

A camada de Aplicações corresponde a uma estrutura onde o usuário poderá pesquisar informações nessa base seja por meio de um formulário simplificado ou por consultas SPARQL realizadas diretamente a um endpoint vinculado a um triplestore onde o grafo estará armazenado.

Com relação à estrutura de operação do FarolKG (processos de carga de dados, extração de entidades, triplificação e população de tabelas), cogita-se implementá-la utilizando a linguagem **Python** [van Rossum 1995] e um banco de dados voltado a grafos como o **Virtuoso** [OpenLink Software 2020], citado anteriormente.

A fase de desenvolvimento do *front-end* ainda não está em discussão, mas este provavelmente deve ser realizado com base em *HTML 5*, *CSS3*, *JavaSscript* e algum *framework Web* a ser definido mais adiante.

4. Metodologia

Este trabalho está dividido nas seguintes fases:

1. Levantamento bibliográfico e teórico, tendo como objetivo o aprofundamento do conhecimento acerca de técnicas de construção de grafos de conhecimento;
2. Pesquisa acerca de ferramentas existentes nesse contexto;
3. Levantamento dos requisitos;
4. Implementação da ferramenta proposta;
5. Realização de testes de usabilidade com usuários não especialistas.
6. Análise dos resultados obtidos nos testes de usabilidade e realização de correções;
7. Publicação dos resultados obtidos em periódicos ou conferências internacionais.

5. Conclusões e Trabalhos Futuros

Estamos propomos neste trabalho a criação de uma ferramenta chamada FarolKG, tendo como finalidade apresentar uma *interface* capaz de permitir aos usuários um ambiente

para consultas a um grafo de conhecimento gerado a partir de uma base de dados de mensagens relacionadas à desinformação.

Um diferencial que pode ser elencado neste trabalho em relação àqueles citados nas referências aos quais está relacionado é que pretende-se trabalhar justamente nas mensagens obtidas do *WhatsApp* ao invés de notícias verificadas, como descrito no ClaimsKG [Tchechmedjiev et al. 2019]. A partir do uso de ferramentas de NLP (Natural Language Processing), algumas informações de interesse do texto das mensagens poderão ser extraídas e atribuídas a novas classes da ontologia de modo a estabelecer relações entre nós do grafo.

Como trabalhos futuros, cogitamos:

1. a expansão de sua base de mensagens para outras mídias sociais;
2. sua integração ao portal **Farol Digital** de modo a ampliar a capacidade deste no contexto de análise de desinformação.

Referências

Cabral, L., Monteiro, J. M., da Silva, J. W. F., Mattos, C. L. C., and Mourão, P. J. C. (2021). Fakewhastapp.br: NLP and machine learning techniques for misinformation detection in brazilian portuguese whatsapp messages. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 63–74. SCITEPRESS.

de Sá, I. C., Monteiro, J. M., da Silva, J. W. F., Medeiros, L. M., Mourão, P. J. C., and da Cunha, L. C. C. (2021). Digital lighthouse: A platform for monitoring public groups in whatsapp. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 297–304. SCITEPRESS.

Gómez-Pérez, J. M., Pan, J. Z., Vetere, G., and Wu, H. (2017). Enterprise knowledge graph: An introduction. In Pan, J. Z., Vetere, G., Gómez-Pérez, J. M., and Wu, H., editors, *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, pages 1–14. Springer.

Jabardi, M. H. and Hadi, A. S. (2020). Twitter Fake Account Detection and Classification using Ontological Engineering and Semantic Web Rule Language. *Karbala International Journal of Modern Science*, 6.

Jabardi, M. H. and Hadi, A. S. (2021). Ontology Meter for Twitter Fake Accounts Detection. *International Journal of Intelligent Engineering and Systems*, 14(1).

Martins, A. D. F., Cabral, L., Mourão, P. J. C., Monteiro, J. M., and Machado, J. C. (2021). Detection of misinformation about COVID-19 in brazilian portuguese whatsapp messages. In Métais, E., Meziane, F., Horacek, H., and Kapetanios, E., editors, *Natural Language Processing and Information Systems - 26th International Conference on Applications of Natural Language to Information Systems, NLDB 2021, Saarbrücken, Germany, June 23-25, 2021, Proceedings*, volume 12801 of *Lecture Notes in Computer Science*, pages 199–206. Springer.

OpenLink Software (2020). Virtuoso. Technical report, OpenLink Software.

Pan, J. Z., Pavlova, S., Li, C., Li, N., Li, Y., and Liu, J. (2018). Content based fake news detection using knowledge graphs. In Vrandecic, D., Bontcheva, K., Suárez-Figueroa, M. C., Presutti, V., Celino, I., Sabou, M., Kaffee, L., and Simperl, E., editors, *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part I*, volume 11136 of *Lecture Notes in Computer Science*, pages 669–683. Springer.

Rehm, G., Schneider, J. M., and Bourgonje, P. (2018). Automatic and manual web annotations in an infrastructure to handle fake news and other online media phenomena. In Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., and Tokunaga, T., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*. European Language Resources Association (ELRA).

Stanford University (2021). Protégé. Technical report, Stanford University.

Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., Zloch, M., Zapilko, B., Dietze, S., and Todorov, K. (2019). Claimskg: A knowledge graph of fact-checked claims. In Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I. F., Hogan, A., Song, J., Lefrançois, M., and Gandon, F., editors, *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II*, volume 11779 of *Lecture Notes in Computer Science*, pages 309–324. Springer.

van Rossum, G. (1995). Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

W3C (2013). SPARQL. Technical report, World Wide Web Consortium (W3C).

Yan, J., Wang, C., Cheng, W., Gao, M., and Zhou, A. (2018). A retrospective of knowledge graphs. *Frontiers Comput. Sci.*, 12(1):55–74.

Zhou, L. and Zhang, D. (2004). Building a misinformation ontology. In *2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004), 20-24 September 2004, Beijing, China*, pages 445–448. IEEE Computer Society.

Zhou, L. and Zhang, D. (2007). An ontology-supported misinformation model: Toward a digital misinformation library. *IEEE Trans. Syst. Man Cybern. Part A*, 37(5):804–813.