FakeKG: A Knowledge Graph of Fake Claims for Improving Automated Fact-Checking (Student Abstract)

Gautam Kishore Shahi

University of Duisburg-Essen, Germany
gautam.shahi@uni-due.de

Abstract

False information could be dangerous if the claim is not debunked timely. Fact-checking organisations get a high volume of claims on different topics with immense velocity. The efficiency of the fact-checkers decreases due to 3V problems volume, velocity and variety. Especially during crises or elections, fact-checkers cannot handle user requests to verify the claim. Until now, no real-time curable centralised corpus of fact-checked articles is available. Also, the same claim is fact-checked by multiple fact-checking organisations with or without judgement. To fill this gap, we introduce FakeKG: A Knowledge Graph-Based approach for improving Automated Fact-checking. FakeKG is a centralised knowledge graph containing fact-checked articles from different sources that can be queried using the SPARQL endpoint. The proposed FakeKG can prescreen claim requests and filter them if the claim is already fact-checked and provide a judgement to the claim. It will also categorise the claim's domain so that the fact-checker can prioritise checking the incoming claims into different groups like health and election. This study proposes an approach for creating FakeKG and its future application for mitigating misinformation.

Introduction

Fake news is a significant problem for our society because it may manipulate or change opinions by spreading misleading content or, in some cases, may lead to worsening situations. For instance, during the COVID-19 pandemic, misleading news created chaos and uncertainty among people. This instance is a prime example of how on-time and reliable fact-checking approaches are necessary to limit the problems of the spread of misleading information.

Predominant approaches to mitigating misinformation can be grouped into two categories. First, humans as experts (fact-checkers) work for fact-checking organisations (e.g., snopes.com, politifact.com). Second, crowd-based fact-checking websites like Twitter Birdwatch (an initiative started in the USA). However, both methods suffer from many drawbacks. While experts classify misinformation fairly accurately, this strategy is difficult to scale due to the limited number of available human experts. Moreover, crowd-based fact-checking requires domain knowledge and background information to debunk claims. However, in certain cases, unifying results might be difficult due to contradicting biases of the community members.

In real-world scenarios, vast requests for fact-checking arise during crises or elections. During crises or elections, fact-checking platforms face the 3V problem, volume: a large number of requests to check for fake news; velocity: during the peak, the speed of getting requests also increases; variety: multimodal data like images, text, and videos are part of the requested fact-checking. Still, fake news detection requires a manual effort to verify the claims. Sometimes a verification process also needs to take advice from experts.

Google started an initiative of ClaimReview[1] and Fact Check Markup Tool[2] to provide structured fact-checked articles. But, it has several errors and contains only limited information. Currently, no real-time centralised corpus of previously fact-checked articles is available. These problems can be addressed by automatic fact-checking (AFC) using semantic technologies. The semantic aware AFC helps to debunk the claims using online resources and detection of fake news and topic detection so that fact-checkers can prioritise fact-checking requests. Influenced by the above problems, in this study, we propose the research question (RQ): How can we improve Automated Fact Checking using Knowledge Graph?

The main contribution of this work is to highlight the issues in the current fact-checking processes and to propose a semantic-technology-based scalable AFC solution, FakeKG: A Knowledge Graph of Fact-Checked News. Unlike ClaimKG (Tchechmedjiev et al. 2019), we also include the claims from social media in the KG. Ultimately, it only asks an expert fact-checker to debunk the claim if the background information is unavailable in past claims or online resources, hence reducing the overall workload.

Related Work

Knowledge Graphs (KGs) appear as a representation infrastructure to support the organisation, integration and representation of journalistic content. KGs are used in several domains, like coronavirus knowledge graph (Michel et al. 2020; Reese et al. 2020), patient monitoring (Dutta and De-

[1] https://schema.org/ClaimReview

[2] https://toolbox.google.com/factcheck/explorer

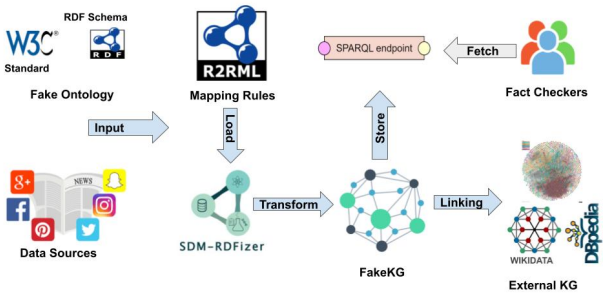

Figure 1: A pipeline for creation of FakeKG

Bellis 2020). The journalism industry lags in data management, data sharing policy, and different structures of data publishing (Shahi and Majchrzak 2022). Pan et al. provided a method for fake news detection using a knowledge graph (Pan et al. 2018). ClaimKG provided metadata, such as truth, and author, which can be used as archival KGs (Tchechmedjiev et al. 2019). So, for the social cause, i.e. fact-checking, there is a need for a central KG which provides on-time access to the structured data so that machines can also process the data, which is accessible in real-time by the different fact-checking organisations and reduces the workload of fact-checkers by finding similarities of fake claims across different languages or claim which is already debunked. This paper proposes that FakeKG helps cross-lingual or multimodal fake news analysis.

Methodology

This section discussed the steps involved in the implementation of FakeKG.

Data Collection: We have collected the previously debunked fake claims and their ratings and stored them in a database. We have collected around 113,006 fact-checked data using the AMUSED framework (Shahi and Majchrzak 2022; Shahi and Nandini 2020) in 40 languages from 105 countries. We stored the collected dataset in a MySQL database.

Knowledge Graph Creation To materialise our idea, we built an ontology (a.k.a Fake Ontology) for creating the FakeKG using the YAMO methodology (Dutta, Madalli, and Chatterjee 2015; Nandini and Shahi 2019). Ontology is used to conceptualise the relational tables' data rows, fields, and relationships into proper classes and properties. We also develop a set of R2RML mapping rules (Sequeda, Priyatna, and Villazón-Terrazas 2012) for each table to transform the table rows to their RDF statement counterparts, systematic and repeatable. The actual work of transforming the table rows to RDF statements is conducted by SDM-RDFizer[3], which reads and applies the mapping rules on the table rows and produces the RDF data. We also link the entity to the external data source using owl:sameAS. In the end, we provide a SPARQL endpoint to query the knowledge, which solves the abovementioned issue. With these queries, the fact-checker can get an answer for complex questions like similar claims or topics, which is difficult to find now from the current unstructured data. For the development of FakeKG, we have a functional pipeline that transforms the collected claims into a knowledge graph, shown in Figure 1.

[3]https://github.com/SDM-TIB/SDM-RDFizer

Conclusion & Future Work

In this paper, we propose a centralised FakeKG for debunking claims. The fact-checker can use the FakeKG as a tool for prescreening claims. The FakeKG will be implemented on data collected from different countries and languages; it will be useful in analysing the propagation of fake news across different news. The tool will be available as open source. In future work, we will implement the proposed approach of FakeKG and provide of list of SPARQL queries as the use case.

References

Dutta, B.; and DeBellis, M. 2020. CODO: An Ontology for Collection and Analysis of Covid-19 Data. *arXiv preprint arXiv:2009.01210*.

Dutta, B.; Madalli, D. P.; and Chatterjee, U. 2015. YAMO: yet another methodology for large-scale faceted ontology construction. *Journal of Knowledge Management*.

Michel, F.; Gandon, F.; Ah-Kane, V.; Bobasheva, A.; Cabrio, E.; Corby, O.; Gazzotti, R.; Giboin, A.; Marro, S.; Mayer, T.; et al. 2020. Covid-on-the-Web: Knowledge Graph and Services to Advance COVID-19 Research. In *International Semantic Web Conference*.

Nandini, D.; and Shahi, G. K. 2019. An ontology for transportation system. *Kalpa Publications in Computing*, 10: 32–37.

Pan, J. Z.; Pavlova, S.; Li, C.; Li, N.; Li, Y.; and Liu, J. 2018. Content based fake news detection using knowledge graphs. In *International semantic web conference*, 669–683. Springer.

Reese, J. T.; Unni, D. R.; Callahan, T. J.; Cappelletti, L.; Ravanmehr, V.; Carbon, S.; Fontana, T.; Blau, H.; Matentzoglu, N.; Harris, N. L.; et al. 2020. KG-COVID-19: a framework to produce customized knowledge graphs for COVID-19 response. *BioRxiv*.

Sequeda, J. F.; Priyatna, F.; and Villazón-Terrazas, B. 2012. Relational Database to RDF Mapping Patterns. In *WOP*.

Shahi, G. K.; and Majchrzak, T. A. 2022. AMUSED: An Annotation Framework of Multimodal Social Media Data. In *International Conference on Intelligent Technologies and Applications*, 287–299. Springer.

Shahi, G. K.; and Nandini, D. 2020. FakeCovid–A Multilingual Cross-domain Fact Check News Dataset for COVID-19. *arXiv preprint arXiv:2006.11343*.

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 *International Semantic Web Conference*, 309–324. Springer.