Knowledge Graphs: Detection of Outdated News

Nicolas Ferranti[1], Astrid Krickl[1], and Markus Nissl[2]

[1] Vienna University of Economics and Business, Vienna, Austria
[2] TU Wien, Vienna, Austria

Abstract. Every day, millions of news stories, articles, and blogs are shared and read throughout the world. As an input, recent study results and everyday occurrences shape how news are released; consequently, it is critical to work with current information, as obsolete information might be incorrect and dangerous in some circumstances, for example, false information about who won the presidential election in the United States, can lead to dangerous public confrontations, as we could observe. In this research, we propose a workflow for the detection of outdated articles with Knowledge Graphs, and analyze the reactivity of updates in DBPedia and Wikidata as possible sources for outdated news.

Keywords: Outdated News · Entity Linking · Knowledge Graphs

1 Introduction

Every day, people are exposed to thousands of news stories containing out-of-date information, such as the current president of a country or official guidelines and limitations during the COVID-19 pandemic. Even books contain outdated information based on obsolete studies and old techniques which are not used anymore. However, when people read such articles, they are unaware and have little possibility of detecting that the information is no longer valid. Given the volume of published literature worldwide, it is apparent that an automatic method for determining the veracity of news stories is essential. The usage of open-source knowledge graphs (KGs) such as DBPedia or Wikidata, which enable centralized storage for structured information, is one option for verifying news articles. In this paper, we present a *workflow* on the integration of DBPedia and Wikidata entries for outdated news detection. For this, we build on state-of-the art retrieval and linking technologies. In addition, we *analyse the contemporaneity* of current DBPedia and Wikidata entries. For this, we study the current information as well as the change history of important personalities.

The remainder of this paper is structured as follows: In Section 2 we provide related work in the area of outdated news detection. In Section 3, we present a workflow for using open source KGs to check for outdated news, and in Section 4 we study and compare the reactivity of open source KGs. Finally, we conclude the paper in Section 5.
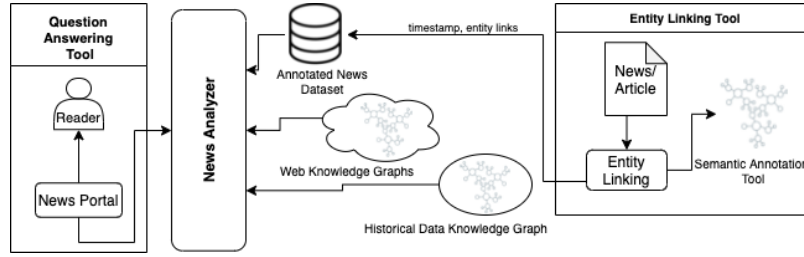
Fig. 1. Workflow to store and analyze semantic annotated news and articles

2   Related Works

Outdated news detection is a type of fake news detection that is generally described as news whose correctness is dependent on a specified time interval. Ongoing research mostly focus on fake news detection [5, 7], in the sense of intentionally and verifiable false [4], where different proposals such as expert fact-checkers, crowdsourcing, natural language processing, or machine learning tools are discussed. In comparison, outdated news require different tools, that are aware of temporal information to detect whether certain facts in articles are still valid. In this context, KGs that store temporal information about the data can be used to validate the news status, as long as they also keep up-to-date facts. Wikidata and DBpedia are among the main cross-domain KGs on the Linked Data Web, although both are based on Wikipedia content, they are very different from each other. David and Guerra et al. [3] present a comparison between Wikidata and DBpedia using four main categories: Intrinsic, Contextual, Representational, and Accessibility. When it comes to timeliness, which is relevant for the outdated news scenario because is based on update rate, the authors show that Wikidata was up to 500 editions/minute while DBpedia is updated about once per month. Frequently updated facts are an important factor in choosing KG, but there are other features that should be taken into account, depending on the purpose of the application. The relationship of KGs and news can be explored in both ways, for different goals. In the literature, there are papers using news to update facts in KGs [9]. Our approach differs from the others using the opposite direction, consuming the KG facts to check the news.

3   Detection of outdated article

We present the workflow for detecting outdated news in Figure 1, which uses KGs to determine whether information in the news is outdated, and then we discuss possible extensions to the basic workflow. The detection process begins with the publication of a new news story, which is automatically fed into an analysis pipeline. Following that, we will outline the pipeline, which is made up of four modules:

Table 1. Comparison between Wikidata and DBpedia of two presidents of the USA

President	Date	Entity and value
Donald Trump DBpedia	2021 June	office:"President of the United States"@en
Donald Trump DBpedia	2018 June	office:"President of the United States"@en
Donald Trump Wikidata	2021 June	Position held (P39): President of the United States
Donald Trump Wikidata	2017 February	Position held (P39): President of the United States
Barack Obama DBpedia	2021 June	office: 44
Barack Obama DBpedia	2015 June	office: 44
Barack Obama DBpedia	2015 June	shortDescription: "American politician, 44th President of the United States"@en
Barack Obama Wikidata	2021 June	Position held (P39): President of the United States
Barack Obama Wikidata	2013 February	Position held (P39): President of the United States

The *entity linking tool* links entities in the news article to KGs such as DB-Pedia. This is performed by using semantic annotation tools, such as DBpedia Spotlight [8] or Dandelion [1] to produce links to the DBPedia ontology which uniquely identify more than 270 resource types, for example, locations, persons, or companies.

An *annotated news dataset* stores the identified links in the article with an associated timestamp of the link detection.

The *news analyzer* (e.g., a deep learning and/or rule-based system) analyzes annotated news article for outdated information by comparing resources of entities from different timestamps. Periodically, or at least when a request rises, the news analyzer extracts the entity links with its corresponding timestamp from the annotated news dataset. Afterwards, it gets the (i) current facts of the KG (for the entity links), and (ii) historical facts at the timestamp from the historical KG (e.g., DBPedia Wayback machine [6]) in order to compare, if something changed.

The *question answering tool* asks the reader of the news if the outdated information was correctly identified. This can be used to train the news analyzer.

As an *example* to this workflow, we use a news article of 2018, which contains information about Donald Trump. The entity linking tool links the entity Donald Trump to the DBPedia entry. When the analyzer analysis the article in 2021, it would receive for example, the following fact for 2018 ⟨Donald_Trump, president, America⟩ in the historical KG, but in 2021 the fact is not present anymore. The reader gets notified that some information about Donald Trump has been changed.

3.1 Extension of the Basic Workflow

In the basic workflow, we compare independent of the article's content the entire information of the identified links. To enhance this approach, we propose three possible improvements:

Contextual Links. Instead of linking resources in the article to the KG, important concepts of the article are extracted as facts. For example, one can extract a fact ⟨*Donald_Trump, president, America*⟩ from the article and link it to the KG.
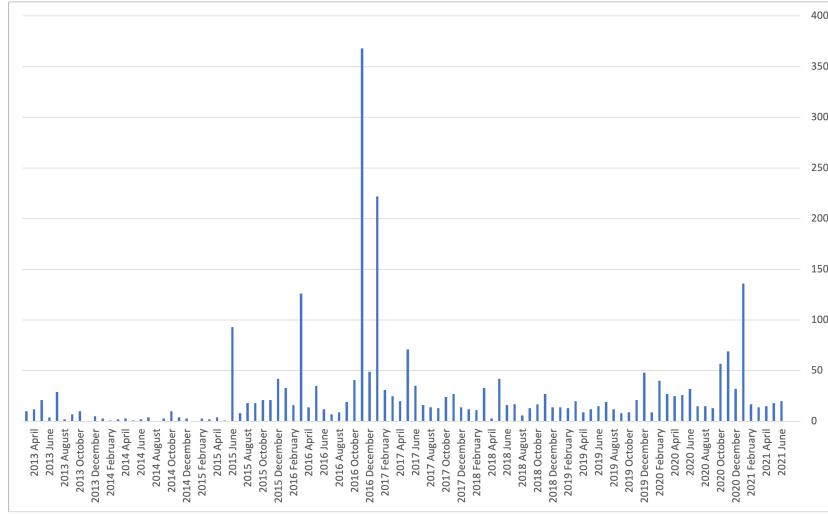
Fig. 2. Number of changes per month for Donald Trump in Wikidata

Comparison of Articles. In addition to KGs, similar contextual links of different articles can be used together with temporal information to extract the currently valid information. This can be used to update either suggest updates to the KGs or be used in the news analyzer to mark articles as outdated.
Post-Processing. Detected changes in the KG can be filtered by checking whether the information is present in the news article. If such change is not relevant, it is discarded and the article is not marked as outdated.

4 Analysis of DBPedia and Wikidata

In this section, we study the reactivity of Wikidata and DBPedia for the verification of outdated articles, and perform some experiments. We analyze knowledge about the topic of the presidents Donald Trump and Barack Obama of the USA to see if there is enough information to verify the claims, as the president of America is one of the most important politician in the world. In order to access current and outdated knowledge we use historical data from Wikidata and the Wayback Machine from DBpedia [6], the results are discussed in Table 1. As it can be seen, the data of Wikidata and DBpedia do not match, however some of the information is not topical. For example, Donald Trump's office says president of the United States which is not correct anymore. Also for some entities the type of information changed in between presidents for example the entity "office" was a number (44) for Barack Obama, as it is a text for Trump. More information about Barack Obama's position can be found in the shortDescription. One entity that got an update on Wikidata is the Position held, as they refer to start dates and later added end dates.

As Table 1 shows, the values are not always clear and up to date. To determine the relevance of the information, we compare the frequency of updates on Wikidata and DBpedia. Because both knowledge graphs use a different method of gathering information, one approach may be better suited to this use case. DBpedia's information originates on Wikipedia and Wikimedia, and while anyone can edit these sources, DBpedia transforms this information and its changes into a DBpedia release multiple times per year [2]. In comparison, Wikidata allows to access the newest version of its Knowledge Graph and has no update cycles. In addition, it is possible to query the history of changes of an entity. We analyzed the frequency of changes for Donald Trump on a monthly basis in Figure 2, which demonstrates that the entity is updated regularly. This is also reflected in Table 1, where we detected an update on the position held property, however updates might be not on time.

5   Conclusion

In this paper, we proposed a workflow for verifying outdated information that makes use of KGs. We described a news analyzer that compares historical KG information with current online KG facts. In order to use our proposal, a up-to-date KG is required. As we demonstrated in the scenario, we currently cannot rely on past data, as it is inaccurate and difficult to compare to fitting entities.

References

1. Entity extraction, https://dandelion.eu/semantic-text/entity-extraction-demo/, [Online; accessed 2021-08-12]
2. FAQ - DBpedia Association, https://www.dbpedia.org/sample-page/, [Online; accessed 2021-08-11]
3. Abián, D., Guerra, F., Martínez-Romanos, J., Trillo-Lado, R.: Wikidata and dbpedia: a comparative study. In: Semanitic Keyword-based Search on Structured Data Sources. pp. 142–154 (2017)
4. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of economic perspectives **31**(2), 211–36 (2017)
5. Collins, B., Hoang, D.T., Nguyen, N.T., Hwang, D.: Fake news types and detection models on social media A state-of-the-art survey. In: ACIIDS (Companion). vol. 1178, pp. 562–573 (2020)
6. Fernández, J.D., Schneider, P., Umbrich, J.: The dbpedia wayback machine. In: SEMANTICS. pp. 192–195 (2015)
7. Meel, P., Vishwakarma, D.K.: Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Syst. Appl. **153**, 112986 (2020)
8. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: Dbpedia spotlight: shedding light on the web of documents. In: I-SEMANTICS. pp. 1–8 (2011)
9. Tang, J., Feng, Y., Zhao, D.: Learning to update knowledge graphs by reading news. In: EMNLP/IJCNLP (1). pp. 2632–2641 (2019)