No Need to Be a Know-It-All: Fact Checking with Shallow Knowledge

Umair Qudus(✉), Neha Pokharel, Michael Röder, and Axel-Cyrille Ngonga Ngomo

DICE Group, Department of Computer Science, Universität Paderborn, Paderborn, Germany
{umair.qudus,neha2022,michael.roeder,
axel.ngonga}@uni-paderborn.de
https://dice-research.org

Abstract. Ensuring the veracity of assertions is vital for building reliable and consistent knowledge graphs. A variety of automatic fact-checking approaches have been proposed over the past decade. Among these, path-based fact-checking approaches are particularly attractive due to their independence of supplementary external knowledge and their faster runtimes compared to methods reliant on external corpora or embeddings. However, the effectiveness of these approaches is fundamentally limited by the incompleteness of existing knowledge graphs, which often lack the paths necessary to support or refute assertions. To address this limitation, we propose ShallKnow, a framework that supplements the knowledge graph with shallow knowledge—automatically extracted RDF assertions from external unstructured sources—even if this additional knowledge may not always fit a well-defined ontology nor be fully verified. By appending such shallow knowledge, we enhance the graph's coverage and increase the chances of finding relevant evidence for fact-checking. Comprehensive experiments on three widely used benchmark datasets demonstrate that integrating ShallKnow consistently and significantly enhances the performance of state-of-the-art path-based fact-checking approaches, yielding improvements of up to 0.24 in Area Under the Receiver Operating Characteristic Curve (AUROC). These results establish ShallKnow as a broadly applicable auxiliary component for improving the reliability and coverage of automatic fact-checking in knowledge graphs. Our code is open-source and can be found at https://github.com/dice-group/ShallKnow.

Keywords: Automatic fact checking · Shallow knowledge · Information extraction · Open Information extraction · Knowledge graphs

1 Introduction

Knowledge graphs (KGs) are essential infrastructure for representing and reasoning over large-scale semantic data on the Web [15,28,32]. As KGs are increasingly used in critical applications such as search engines [50], question answering applications [2], and healthcare systems [35], ensuring the reliability of KGs has become paramount. This is crucial because erroneous or unverifiable assertions within a KG can propagate

M. Acosta et al. (Eds.): ESWC 2026, LNCS 16549, pp. 436–456, 2026.
https://doi.org/10.1007/978-3-032-25156-5_23

misinformation to downstream applications [7,8]. To address this risk, fact checking has emerged as the systematic process of verifying the truth or falsehood of assertions—a task increasingly addressed through automated approaches within KGs [7,27,57]. Consequently, automatic fact checking for KGs has become a vital research challenge, aiming to assess the veracity of assertions and improve the overall trustworthiness and usability of KG-based systems [23,62]. Furthermore, fact checking is pivotal for KG maintenance, as it verifies whether candidate assertions are supported by evidence present within reference knowledge [7,57–59,65]. In response, a variety of automatic fact-checking approaches have emerged, including embedding-based approaches [3,31,44], path-based approaches [34,55,63,66], rule-based approaches [21,22,36,52], text-based approaches [25,65], and hybrid approaches [56,59].

Path-based approaches, in particular, are prominent for their efficiency, scalability, and transparent reasoning, as they leverage only the graph's connectivity and avoid reliance on external data sources or complex model training [56,59]. Nevertheless, these approaches inherently depend on the completeness of the underlying graph.[1] As a result, when key relationships or properties are absent, even sophisticated path-based approaches may fail to find the evidence required to validate or refute a given assertion [60,63].

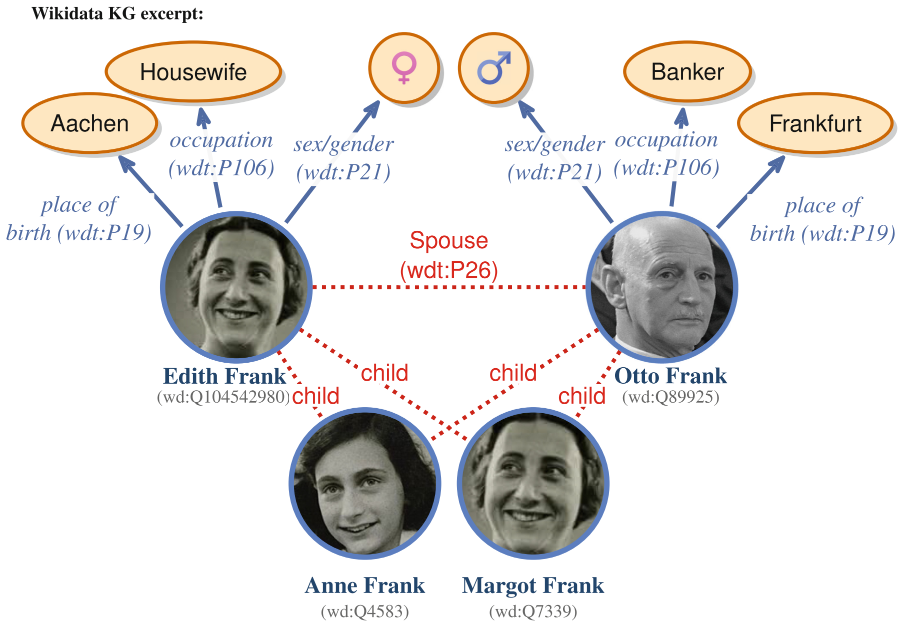
Figure 1 shows an example to illustrate this problem. While entity information such as "Edith Frank" and "Otto Frank" is encoded in the KG, crucial relations like *spouse* or *child* are absent, even though a simple unstructured reference—a snippet from Wikipedia—provides this background knowledge. In such cases, fact-checking approaches restricted to the KG are unable to validate relevant assertions, thereby limiting both coverage and reliability.[2]

A promising direction to overcome this shortcoming is the automatic extraction of "shallow knowledge" from unstructured sources. Shallow knowledge typically takes the form of RDF assertions mined via information extraction, offering broader–albeit less ontologically precise–coverage compared to manual KG curation [39,47]. By appending such assertions, even if noisy or only weakly aligned with KG ontologies, it becomes possible to fill evidential gaps that hinder the fact-checking task. The lower part of Fig. 1 shows this process: The Wikipedia article of "Edith Frank" is analyzed to extract additional relational assertions, which are then added to the KG. These automatically generated assertions, such as missing spouse or child relationships, enhance the KG and allow fact-checking approaches to validate claims that would otherwise lack support from the original KG alone.

In this work, we introduce ShallKnow, a novel, scalable framework for automatic KG fact checking that robustly augments incomplete graphs using shallow structured assertions harvested from large unstructured corpora. Our approach demonstrates that even the integration of imperfect but contextually relevant evidence can substantially increase the success rate of path-based fact checking of assertions. Our work makes the following key contributions:

[1] In this paper, completeness refers strictly to the presence/absence of evidence paths required by path-based fact-checkers, not ontology-level completeness or truth completeness.

[2] We work with IRIs, using the prefixes `wd` and `wdt`. The `wdt` prefix corresponds to the namespace http://www.wikidata.org/prop/direct/, and the `wd` prefix corresponds to the namespace http://www.wikidata.org/entity/.



Wikipedia excerpt:
Edith Frank (1900–1945) and Otto Frank (1889–1980) were married, and had two daughters, Anne Frank (1929–1945) and Margot Betti Frank (1926–1945).
(See: `en.wikipedia.org/wiki/Edith_Frank`*)*

Assertions extracted by ShallKnow:
Primary Triple Extraction:
(`wd:Q104542980, spouse (wdt:P26), wd:Q89925`)
(`wd:Q104542980, child (wdt:P40), wd:Q4583`)
(`wd:Q104542980, child (wdt:P40), wd:Q7339`)
(`wd:Q89925, child (wdt:P40), wd:Q4583`)
(`wd:Q89925, child (wdt:P40), wd:Q7339`)

Secondary Triple Extraction:
(only triggered if the previous step extracted less than θ assertions)
(`wd:Q4583, :P-born-in, wd:Q1794`)
(`wd:Q104542980, :P-married-to, wd:Q89925`)...

Fig. 1. Motivation example. The KG excerpt from Wikidata [46] contains nodes for Edith Frank, Otto Frank, Anne Frank, and Margot Frank, as well as basic attributes, but key familial links (e.g., *spouse*, *child*) are missing (shown as red dotted edges). Wikidata entities mentioned in the figure without full IRIs include: Frankfurt (`wd:Q1794`), Aachen (`wd:Q22094`), Banker (`wd:Q806798`), Housewife (`wd:Q38126150`), Male (`wd:Q6581097`), and Female (`wd:Q6581072`); the relation shown is *child* (`wdt:P40`). θ is a threshold value set by the user.

– We propose ShallKnow, a novel framework for augmenting incomplete KGs with shallow structured assertions automatically extracted from external textual sources.

- We introduce a methodology that leverages large language models (LLMs) for entity-centric text simplification and robust information extraction pipelines for assertion generation.
- We establish an evaluation protocol—including expert manual assessment and inter-annotator agreement analysis—for verifying the factual reliability of shallow knowledge assertions extracted by our system.
- We conduct comprehensive experiments on three state-of-the-art benchmark datasets, demonstrating that supplementing KGs with shallow knowledge significantly improves the performance of multiple path-based fact-checking approaches, yielding AUROC gains of up to 0.24.
- We release our implementation as open source to support reproducibility and further research in the community.[3]

The remainder of this paper is organized as follows. Section 2 introduces the notation and preliminaries needed for the rest of the paper. Section 3 provides an overview of related work. Our proposed approach is presented in Sect. 4, while Sect. 5 describes the experimental setup. Results are discussed in Sect. 6. Finally, Sect. 7 concludes the paper and outlines potential directions for future work.

2 Preliminaries

Our work is concerned with KGs, specifically those represented using the Resource Description Framework (RDF) model [10]. We begin by formally defining an RDF KG.

Definition 1 (RDF KG). *Let E denote the set of all RDF resource IRIs, B the set of blank nodes, $P \subseteq E$ the set of RDF predicates, and L the set of literals. A KG $\mathcal{G}$ is a set of RDF assertions, defined as:*

$$\mathcal{G} = \{(s, p, o) \mid s \in E \cup B,\ p \in P,\ o \in E \cup B \cup L\}, \quad (1)$$

where each assertion (s, p, o) consists of a subject s, predicate p, and object o [10, 55].

The central objective of our approach is the automatic fact checking of KGs. We define the fact-checking task as follows:

Definition 2 (Fact Checking). *Given a candidate assertion in the form of an RDF assertion (s, p, o), together with a reference KG $\mathcal{G}$, fact checking is the process of computing the likelihood with which the assertion is true [65].*

In this work, we introduce the concept of *shallow knowledge* to address the evidence gaps that arise from the incompleteness of KGs [54]. Such knowledge may lack rich ontological structure or comprehensive validation–whether manual or automatic–but can provide additional evidence to support or refute candidate assertions.

[3] We release all code, data, and experimental artifacts through our anonymized GitHub https://github.com/dice-group/shallknow and Zenodo https://zenodo.org/records/15390036 repositories.

Definition 3 (Shallow Knowledge). *Shallow knowledge consists of factual assertions–typically in the form of RDF assertions–automatically extracted from unstructured or semi-structured external textual sources via open knowledge extraction methods [16, 49, 54]. Unlike curated knowledge, shallow knowledge is directly mined from large text corpora without reliance on domain-specific schemas or ontologies.*

3 Related Work

Various automated fact-checking approaches have been introduced in the literature, each leveraging different types of reference data. *Text-based* methods, such as DeFacto [25] and FactCheck [65], cast candidate assertions as search queries to retrieve relevant textual evidence from unstructured corpora. *Path-based* approaches—including Adamic Adar [1], COPAAL [63], Degree Product [61], Jaccard [41], Knowledgestream [61], Katz [33], KL [7], PredPath [60], PRA [24], PathEnt [68], REL-KL [37], and SimRank [30]—analyze evidential patterns, statistical cues, or semantic paths within a KG as the primary information source. These methods aim to capture statistical regularities, express complex semantic dependencies, or detect violations by generalizing from positive and negative examples. Path-based fact-checking approaches are valued for their efficiency and transparency, but their performance is fundamentally constrained by graph incompleteness: such methods often cannot validate or refute assertions when key paths are missing

Recent work, such as FaVEL [55], enriches path selection or systematically combines multiple path-based signals to enhance fact checking performance. Our approach, ShallKnow, further improves the performance of this category by appending shallow knowledge to existing KGs, overcoming limitations due to missing evidence. *Rule- and graph-pattern-based* approaches—such as AMIE [21,22,36], RuDiK [52], GPARs [17], OntoPathFinding [6], and (O)GFCs [42,43]—discover logical rules and frequent subgraph patterns in KGs to infer new facts or validate assertions. While powerful for discovering hidden structure, they rely on sufficient data support and degrade with sparsity or incompleteness. *Embedding-based* models, such as ESTHER [62] and those by Dong et al. [12], rely on KG embeddings to assess assertion plausibility in a latent space. *Hybrid* methods combine multiple evidence modalities—including path-based, rule-based, textual, and embedding-based signals—to improve robustness and fact checking accuracy. Examples include ExFakt [19], Tracy [20], Facty [40], ESTHER [62], HybridFC [59], and TemporalFC [56].

Beyond the KG-based fact-checking methods discussed above, there is extensive research on noisy-knowledge processing [5], triple fusion [14], conflict resolution [45], and trust-aware or ontology-aware KG enrichment [8,13,67,69]. These areas address different challenges, and no prior work has explored enriching KGs with *shallow knowledge* from unstructured text to support fact checking. Our framework does precisely this, providing auxiliary evidence to bridge gaps left by KG incompleteness. Comprehensive evaluation across ten state-of-the-art path-based fact-checking approaches demonstrates that each consistently benefits from the inclusion of shallow knowledge, significantly enhancing both the effectiveness and coverage of path-based fact checking.

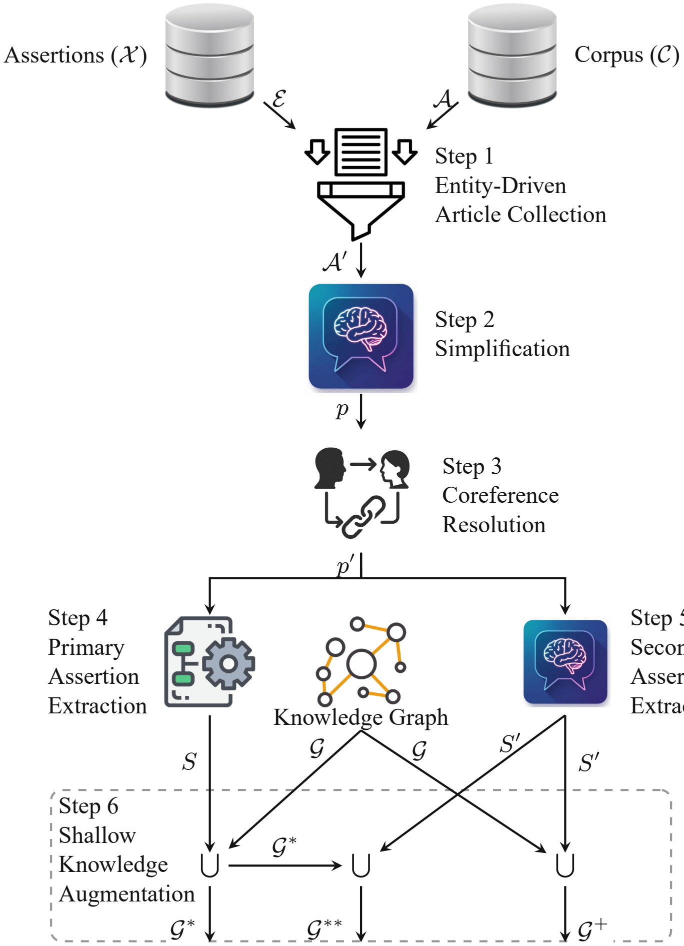


Fig. 2. Shallow knowledge extraction pipeline.

4 Methodology

Our approach systematically augments a KG $\mathcal{G}$ with shallow knowledge assertions $\mathcal{S}$ extracted from a text corpus using a multi-stage pipeline (see Fig. 2 for an overview). This method provides an improved performance on the fact-checking task. In the following, we introduce the problem statement formally.

4.1 Task Formalization

Given a KG $\mathcal{G}$, a set of assertions $\mathcal{X} = \{(s_1, p_1, o_1), \ldots, (s_M, p_M, o_M)\}$ to be verified, and a corpus of unstructured textual articles $\mathcal{C}$, let $\mathcal{E}$ be the set of all entities that occur as subjects or objects in $\mathcal{X}$:

$$\mathcal{E} = \{s_i \mid (s_i, p_i, o_i) \in \mathcal{X}\} \cup \{o_i \mid (s_i, p_i, o_i) \in \mathcal{X}\}. \tag{2}$$

Our objective is to construct three augmented KGs: $\mathcal{G}^* = \mathcal{G} \cup \mathcal{S}$, $\mathcal{G}^+ = \mathcal{G} \cup \mathcal{S}'$, and $\mathcal{G}^{**} = \mathcal{G}^* \cup \mathcal{S}'$, where $\mathcal{S}$ and $\mathcal{S}'$ are defined as:

$$\mathcal{S} = \{(s, p, o) \mid s \in \mathcal{E},\ o \in \mathcal{E},\ p \in \mathcal{P}_\mathcal{G}\}, \tag{3}$$

$$\mathcal{S}' = \{(s, p, o) \mid s \in \mathcal{E},\ o \in \mathcal{E},\ p \in \mathcal{P}_\mathcal{C}\}. \tag{4}$$

Here, $\mathcal{P}_\mathcal{G}$ denotes the set of properties already present in $\mathcal{G}$ (i.e., extracted via KG relation extraction), and $\mathcal{P}_\mathcal{C}$ denotes novel properties or relations mined from the external text corpus $\mathcal{C}$ (e.g., via open knowledge extraction and LLM-assisted relation discovery). The assertions in $\mathcal{S}$ and $\mathcal{S}'$ are extracted by distinct stages of our pipeline, as shown in Fig. 2, and are used to systematically augment the original graph with both established and newly generated properties forming statements about the entities in $\mathcal{E}$.

For example, with reference to Fig. 1, the assertion (`wd:Q89925` [Otto Frank], `wdt:P40` [child], `wd:Q7339` [Margot Frank]) is an instance where $p \in \mathcal{P}_\mathcal{G}$, whereas (`wd:Q4583` [Anne Frank], `:P-born-in` [born in], `wd:Q1794` [Frankfurt]) exemplifies an assertion where $p \in \mathcal{P}_\mathcal{C}$. We next describe the details of our pipeline.

4.2 Entity-Driven Article Collection

For each entity $e \in \mathcal{E}$, we retrieve documents from a reference corpus $\mathcal{C}$ with additional information about these entities. Formally, the initial set of documents is retrieved using a retrieval function q:

$$\mathcal{A} = \bigcup_{e \in \mathcal{E}} q(e)\,. \tag{5}$$

We enrich the context further by expanding this initial set of documents by including all documents that are directly linked from any document $d \in \mathcal{A}$ (i.e., their 1-hop neighbors):

$$\mathcal{A}' = \mathcal{A} \cup \left(\bigcup_{d \in \mathcal{A}} l(d) \right), \tag{6}$$

where l denotes a function retrieving all documents linked from document d. Thus, $\mathcal{A}'$ forms a larger set of relevant documents used for further processing. This strategy increases recall by adding topically related articles to the input set.

4.3 Entity-Centric Paragraph Simplification

Given a document $d \in \mathcal{A}'$, we construct a prompt ζ and apply a large language model (LLM), denoted as f_{LLM}, to process d:

$$p = f_{\text{LLM}}(\zeta, d). \tag{7}$$

Specifically, the LLM is instructed to: 1. Simplify the input text by removing unnecessary tags, markup, and non-informative content; 2. Summarize the remaining information into a clear, plain-language paragraph; 3. Only use information directly present in the text, without generating or inferring new facts; 4. Ensure that, if information about any of the target entities appears in the article, it is preserved in the summary.[4] This procedure produces a concise, entity-centric paragraph p that is free of extraneous details, which makes it suitable for accurate downstream information extraction.

4.4 Coreference Resolution

We perform coreference resolution using a function $\mathcal{R}$ [38]. The idea is to replace pronouns and ambiguous mentions in the paragraph p produced from the previous step with canonical entity names.

$$p' = \mathcal{R}(p). \tag{8}$$

This facilitates accurate downstream assertion extraction [38].

4.5 Primary Assertion Extraction

Our framework supports any assertion extraction tool for generating candidate RDF assertions from a paragraph p'_d summarizing the content of a document d. For our current implementation, we employ REBEL [29] and Relik [51] as extraction modules.[5] Let $\mathcal{T}_k$ denote the extraction function for a tool $k \in \{\text{REBEL}, \text{Relik}\}$. For each $d \in \mathcal{A}'$ and tool k, the extracted assertions are:

$$S_{d,k} = \mathcal{T}_k(p'_d). \tag{9}$$

The union of all such sets across all articles and extraction tools forms the overall primary extracted assertion set:

$$S = \bigcup_{d \in \mathcal{A}} \bigcup_{k} S_{d,k} \,. \tag{10}$$

To exemplify, in the context of Fig. 1, ShallKnow successfully extracts assertions such as (`wd:Q104542980` [Edith Frank], `wdt:P26` [spouse], `wd:Q89925` [Otto Frank]) and (`wd:Q104542980` [Edith Frank], `wdt:P40` [child], `wd:Q4583` [Anne Frank]), illustrating the practical output of this step.

Whenever the total number of assertions extracted by the primary extraction modules for a given paragraph p'_d falls below a user-defined threshold θ, i.e.,

$$\left| \bigcup_{k} S_{d,k} \right| < \theta \,, \tag{11}$$

[4] Due to space constraints, the exact prompts used in our experiments are provided on our project github page for transparency and reproducibility https://github.com/dice-group/ShallKnow/tree/e34783aec974a9d8dc8a7f2ea15dfb8bac0c2792/Prompts.

[5] REBEL and Relik are transformer-based sequence-to-sequence models that perform entity and relation extraction to produce structured assertions directly from input text.

we trigger the secondary assertion extraction as described in the following subsection. This threshold-based mechanism ensures efficient resource usage by invoking the more computationally intensive secondary extraction only when the primary extraction yields comparatively few assertions.

4.6 Secondary Assertion Extraction

If the primary assertion extraction yields fewer than θ assertions for a given processed paragraph p'_d, we apply a secondary extraction step to increase recall. This step consists of two main components.

First, we perform entity recognition to identify candidate named entities that may participate in relational assertions. Specifically, we apply a named entity recognition module, $\mathcal{T}_{\text{NER}}$, to p'_d to obtain a set of entities:

$$\mathcal{E}_d = \mathcal{T}_{\text{NER}}(p'). \tag{12}$$

In our implementation, we use spaCy NER for $\mathcal{T}_{\text{NER}}$.[6] For example, as shown in Fig. 1, recognized entities could include `Edith Frank`, `Otto Frank`, `Anne Frank`, and `Margot Frank`.

Next, we construct a prompt ζ' and use an LLM to extract additional assertions among these entities from p'_d:[7]

$$S'_d = h_{\text{LLM}}(\zeta', p'_d, \mathcal{E}_d), \tag{13}$$

where S'_d denotes the set of secondary assertions inferred for the summary of document d.

To illustrate, in the context of Fig. 1, the secondary assertion extraction produces assertions such as (`wd:Q4583` [Anne Frank], `:P-born-in` [born in], `wd:Q1794` [Frankfurt]) and (`wd:Q104542980` [Edith Frank], `:P-married-to` [married to], `wd:Q89925` [Otto Frank]). These examples demonstrate how this step recovers essential relational information even when standard tools are unable to identify explicit assertions in the text.[8]

However, while the tools used in the previous step rely on the existing ontology of $\mathcal{G}$ and its set of properties $\mathcal{P}_{\mathcal{G}}$, the secondary extraction introduces a set of additional properties $\mathcal{P}_{\mathcal{C}}$.

4.7 Shallow Knowledge Augmentation

Finally, the shallow knowledge is combined with the original KG to obtain $\mathcal{G}^* = \mathcal{G} \cup \mathcal{S}$, $\mathcal{G}^+ = \mathcal{G} \cup S'$, and $\mathcal{G}^{**} = \mathcal{G}^* \cup \mathcal{S}'$. These augmented KGs contain both the original curated assertions and supplementary shallow assertions. Their main difference is that $\mathcal{G}^*$ only comprises properties of $\mathcal{P}_{\mathcal{G}}$, i.e., properties that were already present in

[6] https://spacy.io/api/entityrecognizer.

[7] https://github.com/dice-group/ShallKnow/tree/main/Prompts.

[8] Preliminary experiments with OpenIE methods [16,54] produced limited or low-quality extractions in our KG fact-checking scenario, and thus are not reported in detail here.

the original KG. $\mathcal{G}^+$ comprises additional assertions comprising properties that were newly generated within the secondary assertion extraction. $\mathcal{G}^{**}$ comprises all assertions generated by entire pipeline, i.e., both primary and secondary assertion extraction.

These augmented KGs serve as the basis for performing the fact-checking task in our evaluation. By leveraging both deep and shallow evidence, our approach aims to enhance the accuracy and coverage of automated fact checking.

5 Evaluation

This section describes our evaluation setup. We first delineate the experimental setup and discuss measures undertaken to ensure reproducibility. Subsequently, we outline the fact-checking approaches evaluated in this work. We then present the characteristics of the benchmark datasets employed in our experiments. Finally, we specify the evaluation metric utilized to assess performance.

5.1 Setup and Reproducibility

The goal of our evaluation is to answer whether shallow knowledge can improve the performance of knowledge-graph-based fact-checking approaches. To this end, we use existing fact-checking datasets as input assertions $\mathcal{X}$ and a Wikipedia-based corpus $\mathcal{C}$.[9] For context enrichment, we extract 1-hop neighbor articles for each entity by identifying hyperlinks in the Wikipedia dump and retrieving the corresponding content using the Wikimedia API. Wikidata serves as $\mathcal{G}$ and we execute SHALLKNOW as explained in the previous section to augment it. We define θ as a simple, resource-aware trigger for the secondary extraction step. For an article with $|q|$ sentences, we set $\theta = \left\lceil \frac{2}{3}|q| \right\rceil$. If primary extraction yields fewer assertions than θ, we invoke the secondary step. This proportional threshold balances extraction quality with computational resources, and users may freely adjust it to suit their needs. We then compare the performance of several fact-checking approaches when deployed with $\mathcal{G}, \mathcal{G}^*, \mathcal{G}^+$, or $\mathcal{G}^{**}$.

All experiments are executed on a server with 64 CPU cores, 64 GB RAM, and 1$\times$ NVIDIA RTX 6000 Ada Generation GPU. We use DeepSeek-r1-14B as the large language model (LLM), hosted locally via the Ollama framework.[10] We set the LLM temperature to 0 to enforce greedy decoding, which yields deterministic outputs under standard API settings. DeepSeek-r1-14B [11] is a 14-billion-parameter transformer model trained on both natural language and code tasks. It was selected due to its open licensing, competitive performance on a variety of NLP benchmarks, and ease of deployment for local inference.[11] We use it for the two LLM-based functions f_{LLM} and h_{LLM} in our pipeline.

[9] Any corpus whose documents can be retrieved using entity names can be used instead.

[10] DeepSeek-r1-14B is open source and available via the Ollama framework, which we employ for efficient local setup and inference. More details: https://ollama.com/library/deepseek-r1:14b, https://github.com/deepseek-ai/DeepSeek-V2.

[11] Preliminary experiments with other open-source LLMs (e.g., Mistral, Llama) showed no notable differences, and thus are not reported here due to space constraint.

Runtime vs. θ Tradeoff. The primary extraction requires 0.575 s per document, while the secondary extraction requires 0.383 s. When $\theta = 0$ (secondary extraction disabled), the pipeline processes all documents in 55.9 h. For $\theta = \frac{2}{3}|q|$, secondary extraction is triggered for 77% of documents ($\approx$ 269k), adding 28.7 h and yielding a total runtime of $\approx$ 85.6 hours. Despite the increased coverage, throughput remains comparable (1815.33 vs. 1874.44 triples/hour).

Table 1. Post-processing statistics: numbers of true (T), false (F), and total assertions, and distinct properties (DP) for each dataset.

Dataset	Train (T/F/Total)	Test (T/F/Total)	DP
BPDP 22	100/100/200	103/103/206	2
FactBench Mix 22	633/486/1119	637/492/1129	9
FAVEL-DS	380/385/765	163/164/327	11

5.2 Fact-Checking Approaches

We initially considered the following path-based fact-checking approaches: Katz [33], SimRank [30], Adamic Adar [1], Jaccard [41], Degree Product [61], KL [7], Pathent [68], Knowledgestream [61], PRA [24], PredPath [60], REL-KL [61], COPAAL [63], and FAVEL [55]. Many of these approaches were originally designed for the DBpedia KG and we adapted them to work with Wikidata and our augmented KGs. All approaches except COPAAL and FAVEL are implemented in a single library. As we faced scalability constraints with this library, we modified its codebase to support parallel execution. However, despite substantial effort, three approaches (KL, Knowledgestream, and REL-KL) could not be successfully scaled to larger KGs and were therefore excluded from our final evaluation. For meta-learning within FAVEL, we had to remove the three systems KL, Knowledgestream, and REL-KL and adopted the configuration recommended by [55], utilizing Auto-sklearn 2.0 [18] on top of Scikit-learn [4]. As COPAAL requires ontological information in the form of class subsumption hierarchies and domain and range constraints, we further enriched the Wikidata dump by adding explicit domain and range information. This metadata was extracted by executing SPARQL queries that leverage Wikidata's subject type constraint and value type constraint properties.[12]

5.3 Datasets

For our experiments, we use three benchmark datasets suggested by [55]. These datasets are BPDP [65], FactBench Mix [25], and FAVEL-DS [55], as depicted in Table 1. BPDP 22, FactBench Mix 22 and FAVEL-DS are well-established benchmark datasets that

[12] https://www.wikidata.org/wiki/Q21503250 and https://www.wikidata.org/wiki/Q21510865, respectively. All queries and scripts are publicly available on our GitHub project page.

have been aligned to a DBpedia version of March 2022. To facilitate experiments on the Wikidata KG, we map entities and relations in these datasets to their corresponding Wikidata identifiers using the Wikidata/Wikimedia API[13] and `owl:sameAs` links in Wikidata.

Within our experiments, we use a text corpus ($\mathcal{C}$) derived from the English Wikipedia. We construct this corpus by extracting plain text snippets from all 6.7 million English Wikipedia articles (encompassing approximately 300 million sentences).[14] Since the fact-checking datasets are based on Wikidata, we use the automated entity-to-article mapping of the Wikimedia API for the retrieval function q.[15]

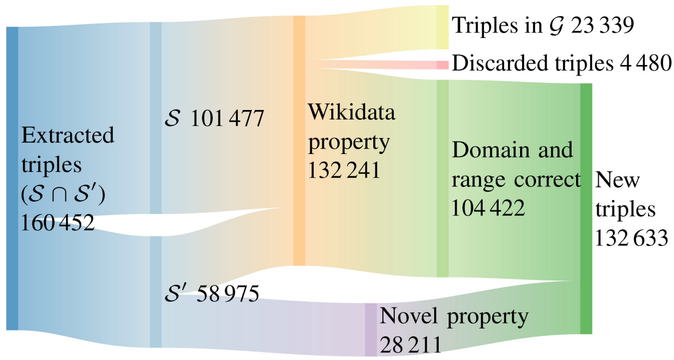


Fig. 3. Sankey diagram illustrating the evaluation of extracted assertions.

5.4 Evaluation Metric

We assess the performance of the fact-checking approaches on the test datasets by measuring the area under the receiver operating characteristic curve (AUROC) as suggested by previous works [34,64,66]. AUROC scores are computed using the GERBIL benchmarking framework [48,53] to ensure consistency and reproducibility of our results.

6 Results and Discussion

To demonstrate the effectiveness of augmenting KGs with shallow knowledge extracted from unstructured corpora, we first present our evaluation methodology for assessing the quality and filtering of extracted assertions. We then provide a comprehensive analysis and discussion of the outcomes from the fact-checking experiments, highlighting both overall performance gains and insights into the impact of shallow knowledge integration.

[13] https://www.wikidata.org/wiki/Help:Sitelinks.
[14] We used the dump from March 7th, 2022 since the corpus should be close to the year of the three benchmarking datasets.
[15] https://api.wikimedia.org/wiki/Main_Page.

6.1 Analysis of Assertion Extraction Results

To assess the quality of shallow knowledge sourced by ShallKnow, we devised a multi-stage evaluation protocol combining automatic schema-based filtering with expert manual assessment, summarized in Fig. 3. For the 3,238 entities in our fact-checking datasets, ShallKnow extracted a total of 160,452 assertions. ShallKnow produced 101,477 assertions through primary extraction ($\mathcal{S}$) and an additional 58,975 assertions via the secondary module ($\mathcal{S}'$). Of the secondary assertions, 28,211 involved novel properties absent from the reference KG schema. Assertions with predicates already in the KG schema or matching existing predicates by verbalization (i.e., `rdfs:label`) underwent further automatic filtering: 1. *Exact Duplicates:* 23,339 assertions matching existing KG assertions were removed. 2. *Invalid Schema Types:* 4,480 assertions were discarded for domain or range violations [54]. We create two sets comprising all schema-aligned assertions passing these checks and all novel-property assertions, respectively. We determine the quality of these assertions by randomly sampling 100 assertions from each set. Two domain experts independently judged factual accuracy of the chosen assertions. The inter-annotator agreement was substantial for novel predicates ($\kappa = 0.6406$) and almost perfect for schema-aligned predicates ($\kappa = 0.8790$) [9]. Discrepancies were resolved via adjudication to come to a final decision. Our pipeline delivers high-quality extractions: novel-property assertions achieved 89% precision, and schema-aligned assertions 86%.

6.2 Fact-Checking Results

Table 2 presents the test set AUROC results for all evaluated approaches and datasets. Augmenting the KG with shallow knowledge via either primary ($\mathcal{G}^*$) or secondary ($\mathcal{G}^+$) assertion extraction consistently improves performance across a broad spectrum of state-of-the-art path-based fact-checking approaches.[16] In almost all scenarios, the combination of assertions from both extraction steps ($\mathcal{G}^{**}$) yields the highest gains, while secondary extraction ($\mathcal{G}^+$) frequently outperforms primary extraction ($\mathcal{G}^*$) alone. For instance, on FaVEL-DS, COPAAL improves from 0.5102 (baseline) to 0.7222 with $\mathcal{G}^*$ ($\Delta = +0.2120$), 0.7400 with $\mathcal{G}^+$ ($\Delta = +0.2298$), and 0.7493 with $\mathcal{G}^{**}$ ($\Delta = +0.2391$). Similarly, PathEnt increases from 0.5733 (baseline) to 0.6221 ($\mathcal{G}^*$, $\Delta = +0.0588$), 0.7400 ($\mathcal{G}^+$, $\Delta = +0.1667$), and 0.7533 ($\mathcal{G}^{**}$, $\Delta = +0.1800$). These results highlight the additive benefit of leveraging both extraction steps, with secondary extraction often providing a greater standalone improvement than primary extraction.

The magnitude of performance improvement depends on both the dataset characteristics and the specifics of the individual method. For example, on the BPDP 22 dataset, which is characterized by relatively well-connected assertions, the gains are generally modest; the highest improvement is observed with PredPath and PRA when utilizing the full pipeline, e.g., Predpath rises from 0.6311 to 0.6748 ($\mathcal{G}^{**}$, Δ=+0.0437) and

[16] $\mathcal{G}$ denotes the original knowledge graph. Let S and S' be the triples extracted by the primary and secondary extraction steps, respectively. We define the augmented graphs as $\mathcal{G}^* = \mathcal{G} \cap S$, $\mathcal{G}^+ = \mathcal{G} \cap S'$, and $\mathcal{G}^{**} = \mathcal{G} \cap S \cap S'$.

Table 2. Test set results of all approaches across all datasets. For each approach, we compare the baseline performance (w/o SHALLKNOW) with results after augmentation using primary assertion extraction ($\mathcal{G}^*$), secondary assertion extraction ($\mathcal{G}^+$), and both extractions combined ($\mathcal{G}^{**}$). Each variant's relative improvement (Δ) is reported. In every dataset (DS), the greatest improvement is shown in **bold**, the second best is underlined, and statistical significance is indicated by a dagger (†).

DS	Approach	w/o SHALLKNOW	SHALLKNOW $\mathcal{G}^*$	Δ	$\mathcal{G}^+$	Δ	$\mathcal{G}^{**}$	Δ
BPDP 22	FAVEL [55]	0.6701	0.6822	†+0.0121	0.7091	†**+0.0390**	0.7232	†+0.0410
	COPAAL [63]	0.5314	0.5497	†<u>+0.0183</u>	0.5370	†+0.0190	0.5515	†+0.0201
	PredPath [60]	0.6311	0.6522	†**+0.0211**	0.6680	†<u>+0.0369</u>	0.6748	†**+0.0437**
	Pathent [68]	0.4985	0.4985	0.0000	0.4988	†+0.0005	0.4990	†+0.0005
	DP [61]	0.5000	0.5000	0.0000	0.5000	0.0000	0.5000	0.0000
	PRA [24]	0.5801	0.5905	†+0.0104	0.6155	†+0.0354	0.6212	†<u>+0.0411</u>
	Jaccard [41]	0.5009	0.5011	†+0.0002	0.5015	†+0.0006	0.5016	†+0.0007
	Adamic Adar [1]	0.4983	0.4983	0.0000	0.4983	0.0000	0.4983	0.0000
	SimRank [30]	0.4817	0.4856	†+0.0039	0.4920	†+0.0087	0.4945	†+0.0100
	Katz [33]	0.4960	0.4963	+0.0003	0.4975	†+0.0015	0.4979	†+0.0017
FactBench Mix 22	FAVEL [55]	0.8133	0.8656	†+0.0223	0.9052	†<u>+0.0919</u>	0.9121	†<u>+0.0988</u>
	COPAAL [63]	0.7903	0.8404	†**+0.0501**	0.8960	†**+0.1057**	0.9011	†**+0.1108**
	PredPath [60]	0.7355	0.7425	+0.0070	0.7766	†+0.0411	0.7895	†+0.0540
	Pathent [68]	0.7054	0.7035	†+0.0019	0.7130	†+0.0076	0.7218	†+0.0164
	DP [61]	0.4467	0.4569	†+0.0102	0.4575	†+0.0108	0.4607	†+0.0140
	PRA [24]	0.5704	0.6033	†<u>+0.0329</u>	0.6170	†+0.0466	0.6221	†+0.0517
	Jaccard [41]	0.6271	0.6344	†+0.0073	0.6537	†+0.0266	0.6621	†+0.0350
	Adamic Adar [1]	0.6591	0.6793	†+0.0202	0.6980	†+0.0389	0.7328	†+0.0737
	SimRank [30]	0.5881	0.6104	†+0.0223	0.6525	†+0.0644	0.6622	†+0.0741
	Katz [33]	0.6002	0.6101	+0.0099	0.6251	†+0.0249	0.6372	†+0.0370
FAVEL-DS	FAVEL [55]	0.7231	0.7532	†+0.0301	0.7831	†+0.0600	0.7921	†+0.0690
	COPAAL [63]	0.5102	0.7222	†**+0.2120**	0.7400	†**+0.2298**	0.7493	†**+0.2391**
	PredPath [60]	0.7189	0.7233	+0.0044	0.7482	†+0.0293	0.7623	†+0.0434
	Pathent [68]	0.5733	0.6221	†<u>+0.0588</u>	0.7400	†<u>+0.1667</u>	0.7533	†<u>+0.1800</u>
	DP [61]	0.5247	0.5247	0.0000	0.5350	†+0.0103	0.5379	†+0.0132
	PRA [24]	0.5322	0.5466	†+0.0144	0.5530	†+0.0208	0.5567	†+0.0245
	Jaccard [41]	0.5398	0.5554	†+0.0156	0.5840	†+0.0442	0.5958	†+0.0560
	Adamic Adar [1]	0.5467	0.5865	†+0.0398	0.6380	†+0.0913	0.6747	†+0.1280
	SimRank [30]	0.5603	0.5778	+0.0175	0.6023	†+0.0420	0.6085	†+0.0428
	Katz [33]	0.5733	0.5735	+0.0002	0.5780	†+0.0047	0.5821	†+0.0880

PRA from 0.5801 to 0.6212 ($\mathcal{G}^{**}$, Δ=+0.0411). This pattern suggests that, while shallow knowledge augmentation is universally beneficial, its greatest impact is realized when the original graph suffers from sparse evidence.

Integration of shallow knowledge also brings notable improvements on FactBench Mix 22, and especially on the FAVEL-DS dataset. On FAVEL-DS, COPAAL [66] achieves the largest gain across all configurations, underscoring the potential of the proposed approach and the sensitivity of advanced path-based algorithms to the introduction of new evidential paths, i.e., COPAAL achieves an AUROC jump from 0.5102 (baseline) to 0.7493 ($\mathcal{G}^{**}$, $\Delta = +0.2391$). PathEnt [68] also shows significant improvement, highlighting that approaches leveraging graph structural features derive substantial benefit from an enriched knowledge base. PathEnt improves from 0.5733 to 0.7533 (+0.1800). This trend is confirmed on FactBench, where COPAAL and FaVeL achieve marked gains, but even simple baselines such as Adamic Adar [1] and SimRank [30] consistently benefit, albeit to a lesser extent. Collectively, these results demonstrate the efficacy of shallow knowledge in bridging gaps in otherwise incomplete paths, thereby enabling both sophisticated and simple metrics to take fuller advantage of the underlying graph.

To gain deeper insight into the impact of SHALLKNOW, we performed a predicate-wise analysis of the predicted scores for both true and false assertions across all datasets (see Fig. 4). Across the majority of predicates, incorporating shallow knowledge leads to higher predicted scores for true assertions, demonstrating that the additional evidence effectively reinforces correct assertions in the KG. This effect is particularly evident for predicates with lower inherent connectivity, where shallow assertions expand the available evidential paths. Conversely, for a minority of predicates, we observe a modest decrease in scores for false assertions, suggesting that shallow knowledge can increase uncertainty for certain infrequent or ambiguous relations. These findings highlight that while shallow knowledge generally improves the discriminative power of fact-checking models, its benefits seem to be predicate-dependent and especially pronounced for relations with initially sparse support in the graph. It is also evident from our analysis that the impact of SHALLKNOW on negative (false) assertions is less noticeable compared to its effect on true assertions. This result is expected, as finding explicit evidence against false assertions is inherently challenging regardless of the supporting source [26]. Even with the integration of shallow knowledge, KGs and external textual evidence are typically constructed to document positive assertions, making the identification or confirmation of negative assertions a common limitation in current automatic fact-checking scenarios.

To rigorously assess the significance of our results, we conducted the Wilcoxon signed-rank test, comparing the outcomes of each method with and without augmentation.[17] The results confirm that our shallow knowledge integration leads to statistically significant improvements in almost all cases, lending strong support to the generalizability and robustness of our framework. In summary, these findings demonstrate that our method not only consistently boosts the performance of existing path-based fact-checking approaches, but does so in a way that is scalable, modular, and broadly appli-

[17] We use a significance threshold $\alpha = 0.05$.

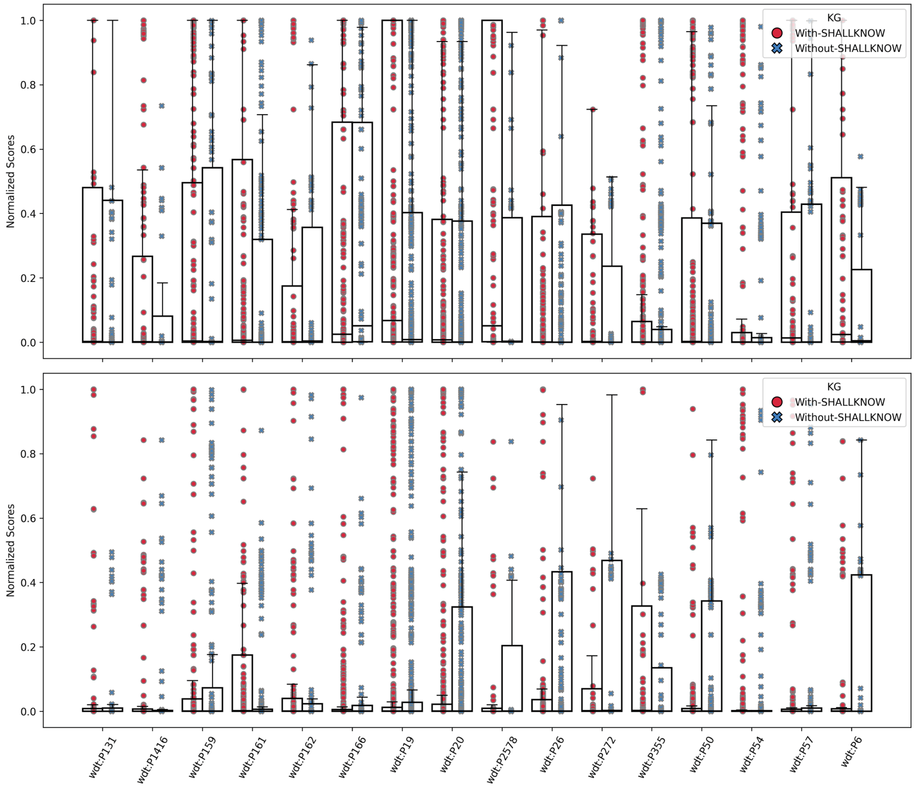


Fig. 4. Predicate-wise predicted scores for true (top, higher values are better) and false (bottom, lower values are better) assertions with and without SHALLKNOW.

cable. This holds significant practical promise for the future of automated validation in large, evolving KGs.

Exploring property normalization and consolidation strategies, as well as mechanisms to better manage rare properties, are promising directions for future work. Moreover, as the extraction process for the full assertion set required approximately 85 h, future efforts may focus on improving computational efficiency and scalability.

7 Conclusion

In this paper, we tackled the dependence of path-based fact-checking approaches on the completeness of supporting evidence within the reference KG. To address this challenge, we introduced a general framework that systematically augments incomplete KGs with shallow knowledge in the form of assertions automatically extracted from unstructured textual sources. Our model-agnostic approach can be applied as an auxiliary evidence enrichment step for various existing fact-checking algorithms. Extensive experiments on three benchmark datasets demonstrate that incorporating shallow knowledge consistently enhances the effectiveness of state-of-the-art path-based fact-checking methods, significantly. The proposed methodology thus bridges critical evidential gaps, contributing to the trustworthiness and reliability of automated KG applications.

Supplemental Material Statement

All supplemental materials are available to ensure full transparency and reproducibility. The source code for SHALLKNOW, all scripts to reproduce our experiments, Docker setup files, prediction results, AUROC graphs, and LLM prompts are available at our anonymized GitHub repository (https://github.com/dice-group/shallknow). The fact-checking datasets used in this study are accessible via Zenodo (https://zenodo.org/records/15390036). Additionally, an Excel file containing all manually evaluated triples with expert annotation results is provided at our anonymized GitHub repository.

Acknowledgments. This work is part of a project that has received funding from the German Federal Ministry of Education and Research (BMBF) within the project NEBULA (13N16364), the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW) within the project SAIL (NW21-059D), and the German Federal Ministry of Research, Technology and Space (BMFTR) within the project KI-Akademie OWL under the grant no 16IS24057B and project Learn2RAG under the grant no. 01MK250104.

References

1. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Social Netw. **25**(3), 211–230 (2003). https://doi.org/10.1016/S0378-8733(03)00009-1
2. Athreya, R.G., Ngonga Ngomo, A.C., Usbeck, R.: Enhancing community interactions with data-driven chatbots–the DBpedia chatbot. In: Companion Proceedings of World Wide Web, pp. 143–146. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018). https://doi.org/10.1145/3184558.3186964
3. Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS, pp. 2787–2795. Curran Associates Inc., Red Hook, NY, USA (2013)
4. Buitinck, L., et al.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122 (2013)
5. Chekol, M.W., Pirrò, G., Schoenfisch, J., Stuckenschmidt, H.: Marrying uncertainty and time in knowledge graphs. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 88–94. AAAI'17, AAAI Press (2017)
6. Chen, Y., Goldberg, S., Wang, D.Z., Johri, S.S.: Ontological pathfinding: mining first-order knowledge from large knowledge bases. In: ICMD, pp. 835–846. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2882903.2882954
7. Ciampaglia, G.L., Shiralkar, P., Rocha, L.M., Bollen, J., Menczer, F., Flammini, A.: Computational fact checking from knowledge networks. PLoS ONE **10**(6), e0128193 (2015)

8. Cimiano, P., Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. Semant. Web **8**(3), 489–508 (2017). https://doi.org/10.3233/SW-160218
9. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. **20**(1), 37–46 (1960). https://doi.org/10.1177/001316446002000104
10. Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1 concepts and abstract syntax. W3C Recommendation, W3C (2014). http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
11. DeepSeek: Deepseek-v1.5-14b model card (2024). https://github.com/deepseek-ai/DeepSeek-V2. Accessed 12 June 2024
12. Dong, T., Wang, Z., Li, J., Bauckhage, C., Cremers, A.B.: Triple classification using regions and fine-grained entity typing. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 77–85 (2019). https://doi.org/10.1609/aaai.v33i01.330177, https://ojs.aaai.org/index.php/AAAI/article/view/3771
13. Dong, X., et al.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: KDD 2014, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 601–610. Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2623330.2623623
14. Dong, X.L., Naumann, F.: Data fusion: resolving data conflicts for integration. Proc. VLDB Endow. **2**(2), 1654–1655 (2009). https://doi.org/10.14778/1687553.1687620
15. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. In: International Conference on Semantic Systems (2016). https://api.semanticscholar.org/CorpusID:8536105
16. Fader, A., Soderland, S., Etzioni, O.: Identifying relations for open information extraction. In: EMNLP ’11, Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1535–1545. Association for Computational Linguistics, USA (2011)
17. Fan, W., Wang, X., Wu, Y., Xu, J.: Association rules with graph patterns. Proc. VLDB Endow. **8**(12), 1502–1513 (2015). https://doi.org/10.14778/2824032.2824048
18. Feurer, M., Eggensperger, K., Falkner, S., Lindauer, M., Hutter, F.: Auto-sklearn 2.0: hands-free AutoML via meta-learning. J. Mach. Learn. Res. **23**(1) (2022)
19. Gad-Elrab, M.H., Stepanova, D., Urbani, J., Weikum, G.: Exfakt: a framework for explaining facts over knowledge graphs and text. In: WSDM ’19, WSDM, pp. 87–95. ACM, New York, NY, USA (2019). https://doi.org/10.1145/3289600.3290996
20. Gad-Elrab, M.H., Stepanova, D., Urbani, J., Weikum, G.: Tracy: tracing facts over knowledge graphs and text. In: WWW ’19, The World Wide Web Conference, pp. 3516–3520. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3308558.3314126
21. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with amie+. VLDB J. **24**(6), 707–730 (2015). https://doi.org/10.1007/s00778-015-0394-1
22. Galárraga, L.A., Teflioudi, C., Hose, K., Suchanek, F.: Amie: Association rule mining under incomplete evidence in ontological knowledge bases. In: World Wide Web ’13, World Wide Web, pp. 413–422. ACM, New York, NY, USA (2013). https://doi.org/10.1145/2488388.2488425
23. Galárraga, L., Razniewski, S., Amarilli, A., Suchanek, F.M.: Predicting completeness in knowledge bases. In: WSDM 2017, Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM (2017). https://doi.org/10.1145/3018661.3018739
24. Gardner, M., Talukdar, P., Krishnamurthy, J., Mitchell, T.: Incorporating vector space similarity in random walk inference over knowledge bases. In: EMNLP, pp. 397–406. Association for Computational Linguistics, Doha, Qatar (2014). https://doi.org/10.3115/v1/D14-1044, https://www.aclweb.org/anthology/D14-1044
25. Gerber, D., et al.: Defacto-temporal and multilingual deep fact validation. Web Semant. **35**(P2), 85–101 (2015). https://doi.org/10.1016/j.websem.2015.08.001

26. Glockner, M., Hou, Y., Gurevych, I.: Missing counter-evidence renders NLP fact-checking unrealistic for misinformation. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 5916–5936. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (2022). https://doi.org/10.18653/v1/2022.emnlp-main.397, https://aclanthology.org/2022.emnlp-main.397/
27. Graves, L.: Understanding the promise and limits of automated fact-checking. Reuters Institute for the Study of Journalism (2018)
28. Hogan, A., et al.: Knowledge graphs. ACM Comput. Surv. **54**(4) (2021). https://doi.org/10.1145/3447772
29. Huguet Cabot, P.L., Navigli, R.: REBEL: relation extraction by end-to-end language generation. In: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 2370–2381. Association for Computational Linguistics, Punta Cana, Dominican Republic (2021). https://aclanthology.org/2021.findings-emnlp.204
30. Jeh, G., Widom, J.: Simrank: a measure of structural-context similarity. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2002)
31. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: IJCNLP, pp. 687–696. Association for Computational Linguistics, Beijing, China (2015). https://doi.org/10.3115/v1/P15-1067, https://www.aclweb.org/anthology/P15-1067
32. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: representation, acquisition, and applications. IEEE Trans. Neural Netw. Learn. Syst. **33**(2), 494–514 (2022). https://doi.org/10.1109/TNNLS.2021.3070843
33. Katz, L.: A new status index derived from sociometric analysis. Psychometrika **18** (1953)
34. Kim, J., Choi, K.s.: Unsupervised fact checking by counter-weighted positive and negative evidential paths in a knowledge graph. In: CICLing, pp. 1677–1686. International Committee on Computational Linguistics, Barcelona, Spain (Online) (2020). https://doi.org/10.18653/v1/2020.coling-main.147, https://www.aclweb.org/anthology/2020.coling-main.147
35. Kotonya, N., Toni, F.: Explainable automated fact-checking for public health claims. arXiv preprint arXiv:2010.09926 (2020)
36. Lajus, J., Galárraga, L., Suchanek, F.: Fast and exact rule mining with amie 3. In: The Semantic Web, pp. 36–52. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-49461-2_3
37. Lao, N., Cohen, W.W.: Relational retrieval using a combination of path-constrained random walks. Mach. Learn. **81**(1), 53–67 (2010)
38. Lee, K., He, L., Lewis, M., Zettlemoyer, L.: End-to-end neural coreference resolution. In: Palmer, M., Hwa, R., Riedel, S. (eds.) Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 188–197. Association for Computational Linguistics, Copenhagen, Denmark (2017). https://doi.org/10.18653/v1/D17-1018, https://aclanthology.org/D17-1018/
39. Lehmann, J., et al.: DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. Semant. Web **6**(2), 167–195 (2015). https://doi.org/10.3233/SW-140134
40. Li, F., Dong, X.L., Langen, A., Li, Y.: Knowledge verification for long-tail verticals. Proc. VLDB Endow. **10**(11), 1370–1381 (2017). https://doi.org/10.14778/3137628.3137646
41. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management (2003)
42. Lin, P., Song, Q., Shen, J., Wu, Y.: Discovering graph patterns for fact checking in knowledge graphs. In: Database Systems for Advanced Applications: 23rd International Conference,

DASFAA 2018, Gold Coast, QLD, Australia, May 21–24, 2018, Proceedings, Part I, pp. 783–801. Springer-Verlag, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-91452-7_50

43. Lin, P., Song, Q., Wu, Y., Pi, J.: Discovering patterns for fact checking in knowledge graphs. J. Data Inf. Qual. **11**(3) (2019). https://doi.org/10.1145/3286488
44. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: AAAI, vol. 29 (2015)
45. Liu, W., Liu, J., Duan, H., Zhang, J., Hu, W., Wei, B.: Truthdiscover: resolving object conflicts on massive linked data. In: WWW '17 Companion, Proceedings of the 26th International Conference on World Wide Web Companion, pp. 243–246. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2017). https://doi.org/10.1145/3041021.3054722
46. Malyshev, S., Krötzsch, M., González, L., Gonsior, J., Bielefeldt, A.: Getting the most out of wikidata: semantic technology usage in wikipedia's knowledge graph. In: Vrandečić, D., et al. (eds.) The Semantic Web - ISWC 2018, pp. 376–394. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_23
47. Martinez-Rodriguez, J.L., Hogan, A., Lopez-Arevalo, I.: Information extraction meets the semantic web: a survey. Semant. Web **11**(2), 255–335 (2020). https://doi.org/10.3233/SW-180333
48. Ngonga Ngomo, A.C., Röder, M., Syed, Z.H.: Semantic web challenge 2019. Website (2019). https://github.com/dice-group/semantic-web-challenge.github.io/. Accessed 30 Mar 2022
49. Niklaus, C., Cetto, M., Freitas, A., Handschuh, S.: A survey on open information extraction. In: Bender, E.M., Derczynski, L., Isabelle, P. (eds.) Proceedings of the 27th International Conference on Computational Linguistics, pp. 3866–3878. Association for Computational Linguistics, Santa Fe, New Mexico, USA (2018). https://aclanthology.org/C18-1326/
50. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J.: Industry-scale knowledge graphs: lessons and challenges. Commun. ACM **62**(8), 36–43 (2019). https://doi.org/10.1145/3331166
51. Orlando, R., Huguet Cabot, P.L., Barba, E., Navigli, R.: Retrieve, read and link: fast and accurate entity linking and relation extraction on an academic budget. In: Findings of the Association for Computational Linguistics: ACL 2024. Association for Computational Linguistics, Bangkok, Thailand (2024)
52. Ortona, S., Meduri, V.V., Papotti, P.: Rudik: rule discovery in knowledge bases. Proc. VLDB Endow. **11**(12), 1946–1949 (2018). https://doi.org/10.14778/3229863.3236231
53. Paulheim, H., Ngonga Ngomo, A.C., Bennett, D.: Semantic web challenge 2018. Website (2018). http://iswc2018.semanticweb.org/semantic-web-challenge-2018/index.html. Accessed 22 May 2023
54. Polat, F., Tiddi, I., Groth, P.: Testing prompt engineering methods for knowledge extraction from text. Semant. Web **16**(2), SW–243719 (2025). https://doi.org/10.3233/SW-243719
55. Qudus, U., Pekarou, F.L.T., Silva, A.A.M.d., Röder, M., Ngomo, A.C.N.: Favel: fact validation ensemble learning. In: Knowledge Engineering and Knowledge Management: 24th International Conference, EKAW 2024, Amsterdam, The Netherlands, November 26–28, 2024, Proceedings, pp. 209–225. Springer-Verlag, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-031-77792-9_13
56. Qudus, U., Röder, M., Kirrane, S., Ngomo, A.C.N.: Temporalfc: a temporal fact checking approach over knowledge graphs. In: Payne, T.R., et al. (eds.) The Semantic Web - ISWC 2023, pp. 465–483. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-47240-4_25
57. Qudus, U., Röder, M., Saleem, M., Ngonga Ngomo, A.C.: Fact checking knowledge graphs – a survey. ACM Comput. Surv. (2025). https://doi.org/10.1145/3749838. Accepted

58. Qudus, U., Röder, M., Vollmers, D., Ngonga Ngomo, A.C.: Exprompt: augmenting prompts using examples as modern baseline for stance classification. In: CIKM '24, Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 3994–3999. Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3627673.3679923
59. Qudus, U., Röder, M., Saleem, M., Ngomo, A.C.N.: Hybridfc: a hybrid fact-checking approach for knowledge graphs. In: International Semantic Web Conference, pp. 462–480. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-19433-7_27, https://papers.dice-research.org/2022/ISWC_HybridFC/public.pdf
60. Shi, B., Weninger, T.: Discriminative predicate path mining for fact checking in knowledge graphs. Knowl. Based Syst. **104**, 123–133 (2016). https://doi.org/10.1016/j.knosys.2016.04.015
61. Shiralkar, P., Flammini, A., Menczer, F., Ciampaglia, G.L.: Finding streams in knowledge graphs to support fact checking. In: ICDM, pp. 859–864 (2017). https://doi.org/10.1109/ICDM.2017.105
62. da Silva, A.A.M., Röder, M., Ngomo, A.C.N.: Using compositional embeddings for fact checking. In: International Semantic Web Conference, pp. 270–286. Springer-Verlag, Berlin, Heidelberg (2021). https://doi.org/10.1007/978-3-030-88361-4_16
63. Syed, Z.H., Röder, M., Ngomo, A.N.: Unsupervised discovery of corroborative paths for fact validation. In: International Semantic Web Conference. Lecture Notes in Computer Science, vol. 11778, pp. 630–646. Springer (2019). https://doi.org/10.1007/978-3-030-30793-6_36
64. Syed, Z.H., Röder, M., Ngonga Ngomo, A.C.: Factcheck: Validating RDF triples using textual evidence. In: CIKM '18, CIKM, pp. 1599–1602. ACM, New York, NY, USA (2018). https://doi.org/10.1145/3269206.3269308
65. Syed, Z.H., Röder, M., Ngomo, A.C.N.: FactCheck: validating RDF triples using textual evidence. In: CIKM '18, Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 1599–1602. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3269206.3269308, https://svn.aksw.org/papers/2018/CIKM_FACTCHECK/public.pdf
66. Syed, Z.H., Srivastava, N., Röder, M., Ngomo, A.C.N.: Copaal - an interface for explaining facts using corroborative paths. In: International Semantic Web Conference (2019)
67. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: Discovering and maintaining links on the web of data. In: ISWC '09, Proceedings of the 8th International Semantic Web Conference, pp. 650–665. Springer-Verlag, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04930-9_41
68. Xu, Z., Pu, C., Yang, J.: Link prediction based on path entropy. Phys. A **456**, 294–301 (2016)
69. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: a survey. Semant. Web **7**(1), 63–93 (2015). https://doi.org/10.3233/SW-150175, https://journals.sagepub.com/doi/abs/10.3233/SW-150175