



Towards Knowledge Graphs Validation Through Weighted Knowledge Sources

Knowledge Graphs and Semantic Web - 3rd Iberoamerican Conference and 2nd Indo-American Conference, KGSWC 2021, Proceedings

Huaman, Elwin; Tauqeer, Amar; Fensel, Anna

https://doi.org/10.1007/978-3-030-91305-2_4

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.




This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact openscience.library@wur.nl



Towards Knowledge Graphs Validation Through Weighted Knowledge Sources

Elwin Huaman¹^(✉), Amar Tauqeer¹, and Anna Fensel^{1,2}

¹ Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, Innsbruck, Austria

{elwin.huaman, amar.tauqeer}@sti2.at

² Wageningen Data Competence Center and Consumption and Healthy Lifestyles Chair Group, Wageningen University & Research, Wageningen, The Netherlands

anna.fensel@wur.nl

Abstract. The performance of applications, such as personal assistants and search engines, relies on high-quality knowledge bases, a.k.a. Knowledge Graphs (KGs). To ensure their quality one important task is knowledge validation, which measures the degree to which statements or triples of KGs are semantically correct. KGs inevitably contain incorrect and incomplete statements, which may hinder their adoption in business applications as they are not trustworthy. In this paper, we propose and implement a Validator that computes a confidence score for every triple and instance in KGs. The computed score is based on finding the same instances across different weighted knowledge sources and comparing their features. We evaluate our approach by comparing its results against a baseline validation. Our results suggest that we can validate KGs with an f-measure of at least 75%. Time-wise, the Validator, performed a validation of 2530 instances in 15 min approximately. Furthermore, we give insights and directions toward a better architecture to tackle KG validation.

Keywords: Knowledge graph validation · Knowledge graph curation · Knowledge graph assessment

1 Introduction

Over the last decade, creating and especially maintaining knowledge bases have gained attention, and therefore large knowledge bases, also known as knowledge graphs (KGs) [12], have been created, either automatically (e.g. NELL [4]), semi-automatically (e.g. DBpedia [2]), or through crowdsourcing (e.g. Freebase [3]). Today, open (e.g. Wikidata) and proprietary (e.g. Knowledge Vault) KGs provide information about entities like hotels, places, restaurants, and statements about them, e.g. address, phone number, and website. With the increasing use of KGs in personal assistant and search engine applications, the need to ensure that statements or triples in KGs are correct arises [9, 13, 19]. For example, Google shows the fact (*Gartenhotel Maria Theresia GmbH, phone, 05223*

© Springer Nature Switzerland AG 2021

B. Villazón-Terrazas et al. (Eds.): KGSWC 2021, CCIS 1459, pp. 47–60, 2021.

https://doi.org/10.1007/978-3-030-91305-2_4

563130), which might be wrong because the *phone* number of the *Gartenhotel* is 05223 56313, moreover, there will be cases where the *phone* number is not up-to-date or is missing [14].

To face this challenge, we developed an approach to validate a KG against different knowledge sources. Our approach involves (1) mapping the different knowledge sources to a common schema (e.g. Schema.org¹), (2) instance matching that ensures that we are comparing the same entity across the different knowledge sources, (3) confidence measurement, which computes a confidence score for each triple and instance in the KG, and (4) visualization that offers an interface to interact with. Furthermore, we describe use cases where our approach can be used.

There have been a few approaches proposed to validate KGs. In this paper, we review methods, tools, and benchmarks for knowledge validation. We found out that most of them focus on validating knowledge against the Web or Wikipedia. For example, the approaches measure the degree to which a statement (e.g. *Paris is the capital of France*) is true based on the number of occurrences of the statement in sources such as Wikipedia, websites, and/or textual corpora. In addition, to the best of our knowledge, no studies have investigated how to validate KGs by collecting matched instances from other weighted structured knowledge sources.

In this paper, we propose a weighted approach that validates a KG against a set of weighted knowledge sources, which have different weight (or degree of importance) for different application scenarios. For example, users can define the degree of importance of a knowledge source according to the task at hand. We validate a KG by finding the same instances across different knowledge sources, comparing their features, and scoring them. The score ranges from 0 to 1, which indicates the degree to which an instance is semantically correct for the task at hand.

This paper is structured as follows. Section 2 presents related state-of-the-art methods, tools, and benchmarks. Section 3 describes our validation approach. We evaluate our approach and show its results in Sect. 4. Furthermore, in Sect. 5 we list use cases where our approach may be needed. Finally, we conclude with Sect. 6, providing some remarks and future work plans.

2 Literature Review

Knowledge Validation (KV), a.k.a. fact checking, is the task of assessing how likely a given fact or statement is true or semantically correct [10, 15, 20]. There are currently several state-of-the-art methods and tools available that are suitable for KV. One of the prior works on automating this task focuses on analysing trustworthiness factors of web search results (e.g. the trustworthiness of web pages based on topic majority, which computes the number of pages related to a query) [17]. Another approach is proposed by Yin et al. [27]. Here, the authors

¹ <https://schema.org/>.

define the trustworthiness of a website based on the confidence of facts provided by the website, for instance, they propose an algorithm called *TruthFinder*. Moreover, [5] present Knowledge Vault, which is a probabilistic knowledge base that combines information extraction and machine learning techniques to compute the probability that a statement is correct. The computed score is based on knowledge extracted from the Web and corroborative paths found on Freebase. However, the Web can yield noisy data and prior knowledge bases may be incomplete. Therefore, we propose an approach that not only takes into account the user’s preferences for weighting knowledge sources, but also complements the existing probabilistic approaches.

We surveyed methods for validating statements in KGs and we distinguish them according to the data used by them, as follows: a) *internal* approaches use the knowledge graph itself as input and b) *external* approaches use external data sources (e.g. DBpedia) as input. In the context of this paper, we only consider the approaches that use external knowledge sources for validating statements.

The external approaches use external sources like the DBpedia source to validate a statement. For instance, there are approaches that use websites information [5, 11, 22], Wikipedia pages [7, 18, 23], DBpedia knowledge base [16, 20], and so on. In contrast to other approaches, [16] present an early stage approach that uses DBpedia to find out *sameAs* links, which are followed for retrieving evidence triples in other knowledge sources and [20] uses DBpedia to retrieve temporal constraints for a fact. However, [16] do not provide an evaluation of the approach to be compared with our approach and [20] focus on validating dynamic data, which we do not tackle in the scope of this paper. Furthermore, there are methods that use topic coherence [1] and information extraction [22] techniques to validate knowledge. Obviously, there is not only one approach or ideal solution to validate KGs. The proposed tools – DeFacto², Leopard³, FactCheck⁴, and FacTify⁵– rely on the Web and/or external knowledge sources like Wikipedia.

The current Web-based approaches can effectively validate knowledge that is well disseminated on the Web, e.g. *Albert Einstein’s date of birth is March 14, 1879*. Furthermore, the confidence score is based on the number of occurrences of a statement in a corpus (e.g. Wikipedia). Unfortunately these approaches are also prone to spamming [24]. Therefore, a new approach is necessary to further improve KG validation. In this paper, we propose a KG validation approach, which computes a confidence score for each triple and instance of KGs.

Furthermore, an evaluation of validation approaches is really important, therefore, we also surveyed knowledge validation benchmarks that have been proposed, however, the number of them is currently rather limited. [26] and [7] released a benchmark consisting of triples extracted from a KG (e.g. Yago) and textual evidences retrieved from a corpus (e.g., Wikipedia). Furthermore. [25]

² <https://github.com/DeFacto/DeFacto>.

³ <https://github.com/dice-group/Leopard>.

⁴ <https://github.com/dice-group/FactCheck>.

⁵ <http://qweb.cs.aau.dk/factify/>.

released FEVER⁶ that is a dataset containing 185K claims about entities which were verified using Wikipedia articles. Moreover, FactBench⁷ (Fact Validation Benchmark) provides a multilingual (i.e. English, German and French) benchmark that describes several relations (e.g. Award, Birth, Death, Foundation Place) of entities.

All benchmarks mentioned above have focused mostly on textual sources, i.e. unstructured information. Therefore, from the best of our knowledge, there is no available benchmark that can be used for validating knowledge graphs via collecting matched instances from other structured knowledge sources.

Last but not least, the reviewed approaches are mostly focused on validating well disseminated knowledge than factual knowledge. Furthermore, benchmarks are built for validating specific tools or to be used during contests like FEVER. Another interesting observation is that Wikipedia is the most frequently used by external approaches (i.e. Wikipedia as textual corpus for finding evidences). Finally, to make future works on knowledge graph validation comparable, it would be useful to have a common selection of benchmarks.

3 Approach

In this section, we present the conceptualization of our KG validation approach. First, we give an overview of the knowledge validation process (see Fig. 1). Second, we state the input needed for our approach in Sect. 3.1. In Sect. 3.2, we describe the need for a common attribute space between knowledge sources. Then, in Sect. 3.3, we explain the instance matching process. Afterwards, confidence measurement of instances is detailed in Sect. 3.4. Finally, in Sect. 3.5, we describe the output of our implemented approach.

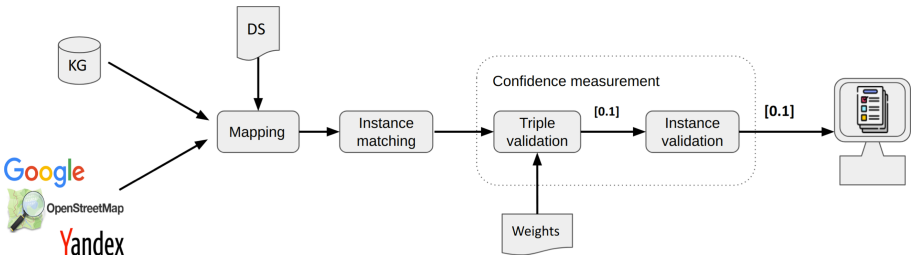


Fig. 1. Knowledge graph validation process overview.

The input to the Validator is a KG, which can be provided via a SPARQL endpoint or an RDF dataset in Turtle⁸ format. This input KG is first mapped

⁶ <https://github.com/sheffieldnlp/fever-naacl-2018>.

⁷ <https://github.com/DeFacto/FactBench>.

⁸ <https://www.w3.org/TR/turtle/>.

based on a Domain Specification⁹ (DS), which basically defines the mapping of the KG to a common format, e.g., this process may be performed by a domain expert, who defines the types and properties that are relevant to the task at hand or the user’s need [21]. A DS defines the instance type and properties values to be validated. Internally, the Validator is configured to retrieve data from external sources, which are also mapped to the common format. After the mapping process has been done, the instance matching is used to find the same instances across the KG and the external sources. Then, the confidence measurement process is triggered and the features of same instances are compared with each other. For example, we compare the name value of an instance of the KG against the name value of the same instance in an external source. We repeat this process for every triple of an instance and we compute a triple confidence score, the triple confidence scores are later added to an aggregated confidence score for the instance. The computed scores are normalized according to the weights given to each knowledge source. We consider the quality of the external sources subjective, therefore, we provide a graphical user interface that allows users to weight each knowledge source.

3.1 Input

At first step, a user is required to provide a KG to be validated. For this, the user has two options, a) to provide a SPARQL endpoint where to fetch the data from or b) to load a dataset in a Turtle format. Moreover, the user is required to select, from a list of DSs, a DS that defines an instance type (e.g., Hotel, Person) and their corresponding properties (e.g., name, address). Internally, the Validator has been set up to fetch data from different external sources (e.g. Wikidata, DBpedia), which were selected based on their domain coverage for the task at hand and their widely use [8].

3.2 Mapping

Based on the DS defined in the input, the validator maps the input KG and the external sources to a common format, e.g., a telephone number of a hotel can be stored with different property names across the knowledge sources: *phone*, *telephone*, or *phone_number*. The validator provides a basic mapping feature to map the input KG and external data sources to a common attribute space. This step is not trivial. There is a huge number of knowledge sources and their schemas might be constantly changing [6]. As a result, schema alignment¹⁰ is one of the major bottlenecks in the mapping process. Therefore, new methods and frameworks to tackle the schema heterogeneity are needed.

⁹ Domain Specification are design patterns for annotating data based on Schema.org. This process implies to remove types and properties from Schema.org, or add types and properties defined in an external extension of Schema.org.

¹⁰ Schema alignment is the task of determining the correspondences between various schemas.

3.3 Instance Matching

So far, we mapped knowledge sources to a common attribute space. However, a major challenge is to match instances across these knowledge sources. For that, the Validator requests to define at least two or more properties (e.g., *name* and *geo coordinates*) that are to be used for the instance matching process, which is constrained to strict matches on the defined property values. The resulting matched instance is returned to the Validator and processed to measure its confidence.

3.4 Confidence Measurement

Computing a confidence value can get complicated as the number of instances and their features can get out of hand quickly. Therefore, a means to automatically validate KGs is desirable. To compute a confidence value for an instance, the confidence value for each of its triples has to be evaluated first.

Triple Validation. calculates a confidence score of whether a property value on various external sources matches the property value in the user’s KG. For example, the user’s KG contains the *Hotel Alpenhof* instance and statements about it; *Hotel Alpenhof’s phone is +4352878550* and *Hotel Alpenhof’s address is Hintertux 750*. Furthermore, there are other sources, like Google Places, that also contain the *Hotel Alpenhof* instance and assertions about it.

The confidence score of (*Hotel Alpenhof*, *phone*, *+4352878550*) triple is computed by comparing the phone property value *+4352878550* against the same property value of the same instance in Google Places. For that, syntactic similarity matching of the attribute values is used. Then the phone property value is compared against a second knowledge source, and so on. Every similarity comparison returns a confidence value that later is added to an aggregated score for the triple.

We define a set of knowledge sources as S , $S = \{s_1, \dots, s_m\}$, $s_i \in S$ with $1 \leq i \leq m$. The user’s KG g consists of a set of instances that are to be validated against the set of knowledge sources S . A knowledge source s_i consists of a set of instances $E = \{e_1, \dots, e_n\}$, $e_j \in E$ with $1 \leq j \leq n$ and an instance e_j consists of a set of attribute values $P = \{p_1, \dots, p_M\}$, $p_k \in P$ for $1 \leq k \leq M$.

Furthermore, *sim* is a similarity function used to compare attribute pair k for two instances. We compute the similarity of an attribute value of two instances a , b . Where a represents an instance in the user’s KG g , denoted $g(a)$, and b represents an instance in the knowledge source s_i , denoted $s_i(b)$.

$$triple_{confidence}(a_{p_k}, S, sim) = \sum_{i=1}^m sim(g(a_{p_k}), s_i(b_{p_k})) \quad (1)$$

Next, users have to set an external weight for each knowledge source s_i , $W = \{\omega_1, \dots, \omega_m\}$ is a set of weights over the knowledge sources, such as ω_i defines a weight of importance for s_i , $0 \leq i \leq m$, $\omega_i \in W$ with $\omega_i \in [0, 1]$

where 0 is the minimum degree of importance and a value of 1 is the maximum degree. For the sum of weights $w_{sum} = \sum_{i=1}^m \omega_i = 1$ has to hold. We compute the weighted triple confidence as follows:

$$triple_{confidence}(a_{p_k}, S, sim, W) = \frac{1}{w_{sum}} \sum_{i=1}^m sim(g(a_{p_k}), s_i(b_{p_k}))w_i \quad (2)$$

The weighted approach¹¹ aims to model the different degrees of importance of different knowledge sources. None of the parameters can be taken out of their context, thus a default weight has to be given whenever the user does not set weights for an external source. The Validator assigns an equivalent weight for each source: $\omega_i = \frac{1}{m}$.

Instance Validation. computes the aggregated score from the attribute space of an instance. Given an instance a that consists of a set of attribute values $P = \{p_1, \dots, p_M\}$, $p_k \in P$ for $1 \leq k \leq M$:

$$instance_{confidence}(a_{p_k}, S, sim, W) = \frac{1}{M} \sum_{k=1}^M triple_{confidence}(a_{p_k}, S, sim, W) \quad (3)$$

The instance confidence measures the degree to which an instance is correct based on the triple confidence of each of its attributes. The instance confidence score is compared against a threshold¹² $t \in [0, 1]$. If $instance_{confidence} > t$ indicates its degree of correctness.

Knowledge Graph Validator

+ add more

Name ▼

Phone ▼

Address ▼

X

X

X

🔍
search

Weights

Google	0.4
OSM	0.3
Yandex	0.3

Property Name	KG	Google Places	Open Street Map	Yandex Places	Confidence
Name	Hotel Alpenhof	Hotel Alpenhof	Hotel Alpenhof	Alpenbad Hotel Hohenhaus	0.7
Phone	+4352878550	+4352878550	Record Not Found.	+4352878550	0.7
Address	Hintertux 750	Hintertux 750, 6294 Hintertux, Österreich	Record Not Found.	Austria, Tux, Hohenhausgasse 774	0.4
					0.6

Fig. 2. Screenshot of the Validator Web interface.

¹¹ To define weights, a proper quality analysis of the knowledge sources must be carried out [8]. It may assist users in defining degrees of importance for each knowledge source.

¹² The default threshold is defined to 0.5.

3.5 Output

The computed scores for triples and instances are shown in a graphical user interface, see Fig. 2. The interface provides many features: it allows users to select multiple properties (e.g. address, name) to be validated, users can assign weights to external sources, it shows instance information from user’s KG and external sources. For example, the Validator shows information of the *Hotel Alpenhof* instance from all sources. It also shows the triple confidence score for each triple, e.g. the triple confidence for the address property is 0.4, because the address value is confirmed only by Google Places.

Tools and Technologies. We implemented our approach in the Validator tool¹³, which has been implemented in JavaScript¹⁴ for retrieving data remotely, and Bootstrap¹⁵ for the user interface.

4 Evaluation

This section describes the evaluation of our approach. The aim of the experiments is to show a qualitative and quantitative analysis of our approach. The setup used for the evaluation is described in Table 1.

Table 1. Evaluation setup

CPU	RAM	OS
AMD Ryzen 7 pro 4750u (16 Cores)	32 GB	Ubuntu 20.04.2 LTS 64-bit

In Sect. 4.1, we compare the Validator’s validation result against a baseline. Next, we look into the scalability of the Validator in Sect. 4.2.

4.1 Qualitative Evaluation

The qualitative evaluation measures the effectiveness of the Validator based on a baseline validation. To do so, first, we describe a dataset to be used on the quality evaluation of our approach, later on we define a setup for the Validator and execute it. Then, we establish a baseline to compare the result of the Validator.

Hotel Dataset. It was fetched from the Tirol Knowledge Graph¹⁶ (TKG), which contains ~15 Billion statements about hotels, places, and more, of the Tirol region. The data inside the TKG are static (e.g. name, phone number) and dynamic (e.g. availability of rooms, prices) and are based on Schema.org

¹³ <https://github.com/AmarTauqeer/graph-validation>.

¹⁴ <https://developer.mozilla.org/en-US/docs/Web/JavaScript>.

¹⁵ <https://getbootstrap.com/>.

¹⁶ <https://graphdb.sti2.at/sparql>.

annotations, which are collected from different sources such as destination management organizations and geographical information systems. We have created a benchmark dataset of 50 hotel instances¹⁷ fetched from the TKG. We randomly selected 50 hotel instances in order to be able to perform a manual validation of their correctness and establish a baseline. The process of creating the Hotel dataset involved manual checking of the correctness of all instances and their attribute values.

Setup and Execution. First, we set up the Hotel dataset on the Validator. Second, we defined external sources, namely: Google Places¹⁸, OpenStreetMap (OSM)¹⁹, and Yandex Places²⁰. Third, we defined the *Hotel* type and *address*, *name*, and *phone* properties that are used for mapping place instances from external sources. Then, for the instance matching process, we set up the *name* and *geo-coordinates* values to search for places within a specified area. We use the built-in feature provided by the external sources (e.g. Nearby Search for Google places) to search for an instance with the same name within a specific area. Furthermore, weights for the external sources are equally distributed. Finally, we run the validation task.

Baseline. In order to evaluate the results of the Validator, a baseline must be established. Given that no prior validation tool addresses exactly the task at hand, we made a manual validation of the Hotel dataset. We computed the precision, recall, and f-measure that a manual validation would achieve (See Fig. 3). During this evaluation, the 50 hotel instances are manually searched and compared to the results coming from each of the external knowledge sources: Google Places, OSM, and Yandex Places. The compared attributes are the *address*, *name*, and *phone*.

Result. We analyse the result of running the Validator on the Hotel dataset. These results are shown in Fig. 3. On one hand, it shows that the Validator performs almost equally similar as the manual evaluation when it comes to *name* and *phone* properties, on the other hand, the Validator does not perform well on the validation of the *address* property. Moreover, the results suggest that we can validate hotel instances with an f-measure of at least 75% on *address*, *name*, and *phone* properties. To interpret the results of our validation run, we choose precision, recall, and f-measure. Given the results of the Validator run, every validated triple result was classified as True Positive, False Positive, True Negative, or False Negative based on the baseline results.

¹⁷ <https://github.com/AmarTauqeer/graph-validation/tree/master/data>.

¹⁸ <https://developers.google.com/maps/documentation/places/>.

¹⁹ <https://www.openstreetmap.org/>.

²⁰ <https://yandex.com/dev/maps/>.

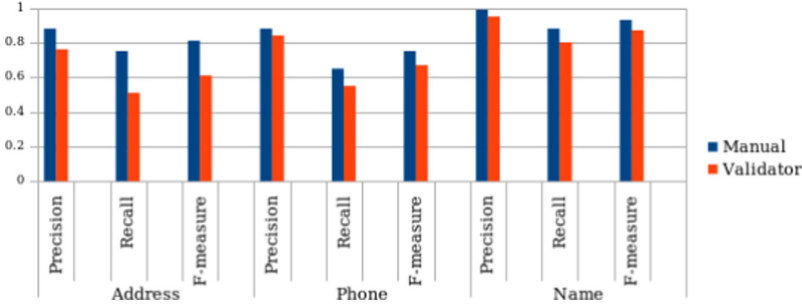


Fig. 3. Comparison of precision, recall, and f-measure scores over the manual and semi-automatic validation.

4.2 Scalability Evaluation

Another challenge of a validation framework is the scalability. In this section, we describe our evaluation approach in terms of scalability of our approach.

Pantheon Dataset. It contains manually validated data with 11341 famous biographies [28]. Pantheon describes information like name, year of birth, place of birth, occupation, and many more. We have selected politician domain and created a dataset of 2530 politician instances. We selected the politician domain because it has the highest number of instances in the Pantheon dataset. Furthermore, we had to convert the Pantheon dataset to Turtle format, for that we used Tarql²¹ tool. Last but not least, we selected the politician domain in order to prove the general applicability of our approach in different domains (e.g., Hotel, Person).

Setup and Execution. The setup for validating datasets from different domains changes slightly, for example, defining the external sources where to fetch the data from. First, we set up the Pantheon dataset on the Validator. Then, we defined Wikidata and DBpedia as external sources and we distributed equivalent weights for them. Moreover, we defined the *person* type and *name* and *year of birth* properties for mapping politicians from the external sources. Moreover, we set up the *name* and *year of birth* for the instance matching process. Finally, we execute the validation task.

Result. We validated 2530 politician instances by using the Validator, which compares and computes a confidence score for each triple and instance. To execute this task the Validator required ~ 15 min approximately on a CPU described on Table 1. Results are presented in Fig. 4. On one hand, it shows that Wikidata outperforms DBpedia on validated properties, on the other hand, it shows lower

²¹ <https://tarql.github.io/>.

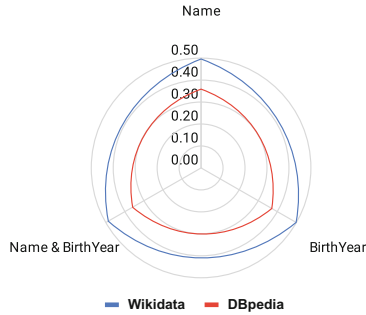


Fig. 4. The recall score results of the validation of politician instances.

recall scores, by the Validator, on both sources, e.g. the overall recall scores are 0.36% (DBpedia) and 0.49% (Wikidata).

Furthermore, the Validator gets lower recall on DBpedia and Wikidata sources due to two reasons. First, DBpedia contains the validated politician instances, however many of them are classified in DBpedia as *agent* type and not as politician (e.g., *Juan Carlos I*²²). Second, the Wikidata query service raised timeout errors when querying data, so we decided to fetch the maximum allowed number of politician instances from Wikidata and stored them locally. We fetched 45000 out of 670810 politicians.

5 Use Cases

Our approach, as described in Sect. 3, aims to validate KGs by finding the same instances across different knowledge sources and comparing their features. Later on, based on the compared features our approach computes a confidence score for each triple and instance, the confidence score ranges from 0 to 1 and indicates the degree to which an instance is correct. Our approach may be used in a variety of use cases, we list some of the cases where the approach can be used:

- To validate the semantic correctness of a triple, e.g., to validate if the phone number of a hotel is the correct based on different sources.
- To link instances between knowledge sources, e.g. linking an instance of the user’s KG with the matched instance in Wikidata.
- To find out incorrect data on different knowledge sources. For instance, suppose that the owner of a hotel wants to validate whether the information of his or her hotel provided by an external source are up-to-date.
- To validate static data, for example, to check whether the addresses of hotels are still valid given a period of time.

There are more possible use cases where our validation approach is applicable. Here, we presented some of them to give an idea about how useful and necessary is to have a validated KG (i.e. a correct and reliable KG).

²² https://dbpedia.org/page/Juan_Carlos.I.

6 Conclusion and Future Work

In this paper, we presented the conceptualization of a new KG validation approach and a first prototypical implementation thereof. Our approach measures the degree to which every instance in a KG is semantically correct. It evaluates the correctness of instances based on external sources. Experiments were conducted on two datasets. The results confirm its effectiveness and are promising great potential. In future work, we will improve our approach and overcome its limitations. Here, we give a short overview of them:

- **Assessment** of knowledge sources. Finding the most suitable knowledge source for validating a KG is challenging [8]. Therefore, it is desirable to implement a quality assessment mechanism for assessing external sources. It may assist users in defining degrees of importance for each knowledge source.
- **Automation** of the setting process. It is desirable to allow users to create a semi-automatic **mapping** (or schema alignment [6]) between their KG and the external sources, e.g. the heterogeneous scheme of OSM has caused low performance of the Validator (see Sect. 4.1).
- **Cost-sensitive methods**. The current version of the Validator relies on proprietary services like Google, which can lead to high costs when validating large KGs. Therefore, it is important to evaluate the cost-effectiveness of knowledge sources.
- **Dynamic data** is fast-changing data that also needs to be validated, e.g. the price of a hotel room. The scope of this paper only comprises the validation of static data.
- **Scalability** is a critical point when we want to validate KGs. KGs are very large semantic networks that can contain billions of statements.

Above, we pointed out some future research directions and improvements that one can implement on the development of future validation tools.

Acknowledgments. This work has been partially funded by the project WordLiftNG within the Eureka, Eurostars Programme of the European Union (grant agreement number 877857 with the Austrian Research Promotion Agency (FFG)) and the industrial research project MindLab (<https://mindlab.ai/>). We would like to thank Prof. Dr. Dieter Fensel for his insightful comments regarding the definition of the overall validation approach.

References

1. Aletras, N., Stevenson, M.: Evaluating topic coherence using distributional semantics. In: Proceedings of the 10th International Conference on Computational Semantics, (IWCS2013), Potsdam, Germany, 19–22 March 2013, pp. 13–22. The Association for Computer Linguistics (2013)

2. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ASWC/ISWC 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52
3. Bollacker, K.D., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM International Conference on Management of Data (SIGMOD2008), Vancouver, Canada, 09–12 June 2008, pp. 1247–1250. ACM (2008)
4. Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E.R., Mitchell, T.M.: Toward an architecture for never-ending language learning. In: Proceedings of the 24th Conference on Artificial Intelligence (AAAI2010), Atlanta, Georgia, 11–15 July 2010, pp. 1306–1313. AAAI Press (2010)
5. Dong, X., et al.: Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining (KDD2014), New York, USA, 24–27 August 2014, pp. 601–610. ACM (2014)
6. Dong, X.L., Srivastava, D.: Big Data Integration. Synthesis Lectures on Data Management, Morgan & Claypool Publishers (2015)
7. Ercan, G., Elbassuoni, S., Hose, K.: Retrieving textual evidence for knowledge graph facts. In: Hitzler, P., et al. (eds.) ESWC 2019. LNCS, vol. 11503, pp. 52–67. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_4
8. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semant. Web* **9**(1), 77–129 (2018)
9. Fensel, D., et al.: Knowledge Graphs - Methodology, Tools and Selected Use Cases. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-37439-6>
10. Gad-Elrab, M.H., Stepanova, D., Urbani, J., Weikum, G.: Tracy: tracing facts over knowledge graphs and text. In: Proceedings of the 19th World Wide Web Conference (WWW2019), San Francisco, USA, 13–17 May 2019, pp. 3516–3520. ACM (2019)
11. Gerber, D., Esteves, D., Lehmann, J., Bühmann, L., Usbeck, R., Ngomo, A.N., Speck, R.: DeFacto - temporal and multilingual deep fact validation. *J. Web Semant.* **35**, 85–101 (2015)
12. Hogan, A., et al.: Knowledge graphs. *ACM Comput. Surv.* **54**(4), 71:1–71:37 (2021)
13. Huaman, E., Kärle, E., Fensel, D.: Knowledge graph validation. *CoRR abs/2005.01389* (2020)
14. Kärle, E., Fensel, A., Toma, I., Fensel, D.: Why are there more hotels in Tyrol than in Austria? Analyzing schema.org usage in the hotel domain. In: Inversini, A., Schegg, R. (eds.) Information and Communication Technologies in Tourism 2016, pp. 99–112. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28231-2_8
15. Lehmann, J., Gerber, D., Morsey, M., Ngonga Ngomo, A.-C.: DeFacto - deep fact validation. In: Cudré-Mauroux, P., et al. (eds.) ISWC 2012. LNCS, vol. 7649, pp. 312–327. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35176-1_20
16. Liu, S., d’Aquin, M., Motta, E.: Towards linked data fact validation through measuring consensus. In: Proceedings of the 2nd Workshop on Linked Data Quality co-located with 12th Extended Semantic Web Conference (ESWC2015), Portorož, Slovenia, 1 June 2015. CEUR Workshop Proceedings, vol. 1376. CEUR-WS.org (2015)
17. Nakamura, S., et al.: Trustworthiness analysis of web search results. In: Kovács, L., Fuhr, N., Meghini, C. (eds.) ECDL 2007. LNCS, vol. 4675, pp. 38–49. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74851-9_4

18. Padia, A., Ferraro, F., Finin, T.: SURFACE: semantically rich fact validation with explanations. CoRR abs/1810.13223 (2018)
19. Paulheim, H.: Knowledge graph refinement: a survey of approaches and evaluation methods. *Semant. Web* **8**(3), 489–508 (2017)
20. Rula, A., et al.: TISCO: temporal scoping of facts. *J. Web Semant.* **54**, 72–86 (2019)
21. Şimşek, U., Angele, K., Kärle, E., Panasiuk, O., Fensel, D.: Domain-specific customization of schema.org based on SHACL. In: Pan, J.Z., et al. (eds.) ISWC 2020. LNCS, vol. 12507, pp. 585–600. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62466-8_36
22. Speck, R., Ngomo, A.N.: Leopard - a baseline approach to attribute prediction and validation for knowledge graph population. *J. Web Semant.* **55**, 102–107 (2019)
23. Syed, Z.H., Röder, M., Ngomo, A.N.: FactCheck: validating RDF triples using textual evidence. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, (CIKM2018), Torino, Italy, 22–26 October 2018, pp. 1599–1602. ACM (2018)
24. Tan, C.H., Agichtein, E., Ipeirotis, P., Gabrilovich, E.: Trust, but verify: predicting contribution quality for knowledge base construction and curation. In: Seventh ACM International Conference on Web Search and Data Mining, WSDM 2014, New York, NY, USA, 24–28 February 2014, pp. 553–562. ACM (2014)
25. Thorne, J., Vlachos, A., Christodoulopoulos, C., Mittal, A.: FEVER: a large-scale dataset for fact extraction and verification. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2018), New Orleans, USA, 1–6 June 2018, pp. 809–819. Association for Computational Linguistics (2018)
26. Vlachos, A., Riedel, S.: Fact checking: task definition and dataset construction. In: Proceedings of the Workshop on Language Technologies and Computational Social Science (ACL2014), Baltimore, USA, 26 June 2014, pp. 18–22. Association for Computational Linguistics (2014)
27. Yin, X., Han, J., Yu, P.S.: Truth discovery with multiple conflicting information providers on the web. *IEEE Trans. Knowl. Data Eng.* **20**(6), 796–808 (2008)
28. Yu, A.Z., Ronen, S., Hu, K., Lu, T., Hidalgo, C.A.: Pantheon 1.0, a manually verified dataset of globally famous biographies. *Sci. Data* **3**(1), 1–16 (2016)