KGHeartBeat: a Knowledge Graph Quality Assessment Tool

Maria Angela Pellegrino[1][0000−0001−8927−5833], Anisa Rula[2][0000−0002−8046−7502], and Gabriele Tuozzo[1]

[1] Dipartimento di Informatica, Unviersità degli Studi di Salerno, Fisciano (SA), ITALY mapellegrino@unisa.it, g.tuozzo4@studenti.unisa.it
[2] Department of Information Engineering, University of Brescia, Brescia, Italy anisa.rula@unibs.it

Abstract. This demo proposes KGHeartBeat, a community-shared open-source knowledge graph quality assessment tool to periodically perform quality analysis on all the freely available knowledge graphs registered on the LOD cloud and DataHub. As a proof of concept, we discuss the comparison of different linguistic versions of DBpedia via KGHeartBeat.

Keywords: Quality assessment · Knowledge Graph · Framework

1 Background and motivation

A considerable amount of data is published using the Semantic Web technologies [8], but they range from extensively curated to relatively low-quality Knowledge Graphs (KGs) [7]. Data quality assessment is a multidimensional problem encompassing heterogeneous and multiple quality dimensions including but not limited to accessibility, interlinking, performance, syntactic validity, and completeness [10]. Several quality assessment tools have been proposed over time, such as RDFUnit [7] (formerly DataBugger), Luzzu [3], SPARQLES [9], SemQuire [8], DYLDO [5], LODLaundromat [2], ABECTO [6]. However, there is no KG quality assessment tool as a reference in the Semantic Web community.

This demo presents KGHeartBeat, a community-shared open-source tool designed to facilitate the assessment and comparison of KGs based on several quality metrics. This tool represents a significant contribution to the field of KG, offering developers and lay users a comprehensive solution for assessing the quality of KGs. While developers are provided with APIs[3] to integrate quality metric computation in any data management workflow, lay users can utilize a user-friendly web-based interface to explore KG quality results visually. The demo primarily focuses on showcasing the KGHeartBeat web application interface[4], which allows users to compare linguistic versions of DBpedia. The interface offers intuitive features for exploring and comparing KG quality metrics.

[3] KGHeartBeat API: https://pypi.org/project/kgheartbeat/
[4] KGHeartBeat web application: http://www.isislab.it:12280/kgheartbeat

2 KGHeartBeat

`KGHeartBeat` is a fully automatic KG quality assessment community-shared framework, publicly available on GitHub[5]. KGHeartBeat weekly computes the quality assessment of all the KGs that can be automatically retrieved by widely used data and knowledge aggregation platforms, e.g., LOD Cloud[6] and DataHub[7]. The metrics computations rely on data retrieved by working SPARQL endpoints, metadata contained in the VoID file and those returned by platforms for data and knowledge aggregation. `KGHeartBeat` implements a large set of well-known quality metrics proposed by Zaveri et al. [10] belonging to different quality dimensions, focusing on those that can be automatically and objectively computed without requiring a gold standard. The implementation details of all the supported quality metrics are freely accessible online[8]. Results can be downloaded as CSV files or visually explored via a freely accessible web application[4] visible in Fig. 1. Users are initially prompted to choose their desired KG(s), after which they can visually explore quality dimensions presented in graphical charts via the web interface.Quality dimensions can be selected from the left-side panel, as shown in Fig. 1 (2), with the corresponding chart displayed in the central panel, as seen in Fig. 1 (4). Quality scores are presented in a simple table format or a more complex chart, depending on the selected quality dimension. Data visualizations aim to enhance understanding for end-users, making assessment and comparison easier to grasp. Quality metric scores can be examined for a specific date, configurable through Fig. 1 (3), or analyzed over time.

Metrics' ratings are then linearly combined into an overall quality assessment score with a numeric value ranging from 0.0 to 100.0, with higher scores indicating better quality. In the KGs ranking tab, users can access the quality scores of all KGs automatically analyzed by `KGHeartBeat`. Moreover, in the `View Score` tab, users can view quality scores specific to the selected KGs. For example, Fig. 2(a) shows the ranking computed for the linguistic versions of DBpedia. In this tab, end-users can customize weights assigned to each metric, allowing them to tailor quality scores to match the use case of interest requirements. Both tabs are accessible via the top-level panel shown in Fig. 1(1).

3 Use Case Driven Metrics and Results

This section overviews metrics and results concerning the Information disorder and automatic fact-checking use case. The University of Salerno is involved in the SERICS project [1], which focuses on security and rights in cyberspace, with one of its key objectives being the detection and mitigation of information disorder, encompassing a wide range of misinformation. Among different perspectives

[5]KGHeartBeat repository: `https://github.com/isislab-unisa/KGHeartbeat`
Permanent URL: `https://zenodo.org/records/10275888`

[6]LOD Cloud: `https://lod-cloud.net`

[7]DataHub: `https://datahub.io`

[8]Metric details: `https://isislab-unisa.github.io/KGHeartbeat`

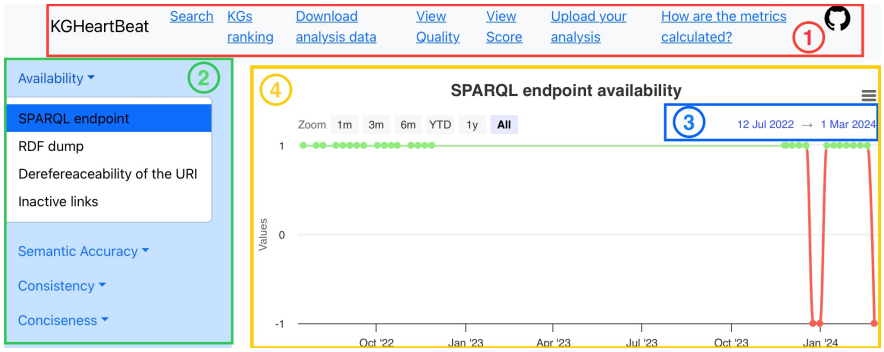


Fig. 1: KGHeartBeat interface. The top-level panel (1) shows the navigation bar, the left-side panel (2) lets users explore quality dimensions, the calendar (3) gives the possibility to customize the time frame of reported quality dimensions scores, and the central panel (4) overviews quality dimensions results graphically according to the end-users configuration.

that can be used to debunk misinformation, KGs play a crucial role when the information content must be explored and automatically compared with external sources [11]. It requires evaluating the trustworthiness and timeliness of information sources [4], emphasizing the credibility of the data and its dynamic nature. To do so, KGHeartBeat can be configured to prioritize the dimensions of trust and dataset dynamicity in computing the final score. The top-5 KGs are reported in Table 1. All of them are curated by (national) organizations, but PGxLOD that is part of the PractiKPharma project[9]. Scores are rather low. Hence, further effort should be invested in curating trust and dynamicity dimensions.

Table 1: Top-5 KGs according to Trust and Data dynamicity dimensions. Scores range from 0 to 100. The higher, the better for all the dimensions. SPARQL stands for SPARQL endpoint.

Metric	**Def**				**Input**	**Output**
Verifiability	Provenance details				SPARQL & VoID	[0,1]
Reputation	Credit computed via the PageRank				Metadata	[0,1]
Believability	The provider is a trustful				Metadata	[0,1]
Currency	Freshness of data [10]				SPARQL & VoID	[0,1]
Timeliness	Presence of frequency of data validation [10]				SPARQL	{0,1}
KG	**Verifiability**	**Reputation**	**Believability**	**Currency**	**Timeliness**	**Score**
Italian Chamber of Deputies	0.66	$8.57e-7$	0.37	1.0	1.0	60.80
Bibliography of the Italian Parliament and electoral studies	0.66	$8.11e-3$	0.25	1.0	1.0	58.30
PGxLOD	0.49	$8.64e-3$	0.37	1.0	1.0	57.47
Corporate Body Named Authority List	0.49	$8.71e-3$	0.37	1.0	1.0	57.47
Country Name Authority List	0.49	$8.61e-4$	0.37	1.0	1.0	57.47

[9] https://practikpharma.mystrikingly.com

4 Demonstration

This section overviews how to use the `KGHeartBeat` web application[4] in practice. Demonstration videos are available in the GitHub repository[10]. Supposing that we are interested in comparing the different linguistic versions of DBpedia as a proof-of-concept. We select all the available linguistic versions of DBpedia, resulting in nine different KGs, listed in Fig. 2(a). Figures 2 and 3 overview some of the quality dimensions scores as graphically rendered by `KGHeartBeat`. The KGs quality is extremely heterogeneous, spanning from 8/100 for the German version of DBpedia to 57/100 for its French version. The quality assessment is heavily impacted by the availability of a working SPARQL endpoint (see Fig. 2(b)). Linguistic versions of DBpedia attached to an offline SPARQL endpoint during the analysis (February 25th, 2024) are ranked as the worst in the overall quality score table visible in Fig. 2(a). `KGHeartBeat` adopts a best-effort approach to compute metrics. When a KG is attached to a working SPARQL endpoint, metric computations rely on current data. As an alternative, it looks for the corresponding value in metadata. The amount of data metric (visible in Fig. 3(a)) is an example in this direction. Fig. 3(b) shows the *Consistency* dimension. As all the KGs reach almost the same score in this dimension, lines are in overlap, and exact values are visible by hovering chart points corresponding to KGs.

KG name	Score
DBpedia in French	56.9
DBpedia	48.2
DBpedia in Spanish	45.7
DBpedia in Japanese	43.4
DBpedia in Dutch	14.6
DBpedia in Greek	13.7
DBpedia in Portuguese	13.1
DBpedia in Basque	13.1
DBpedia in German	7.9

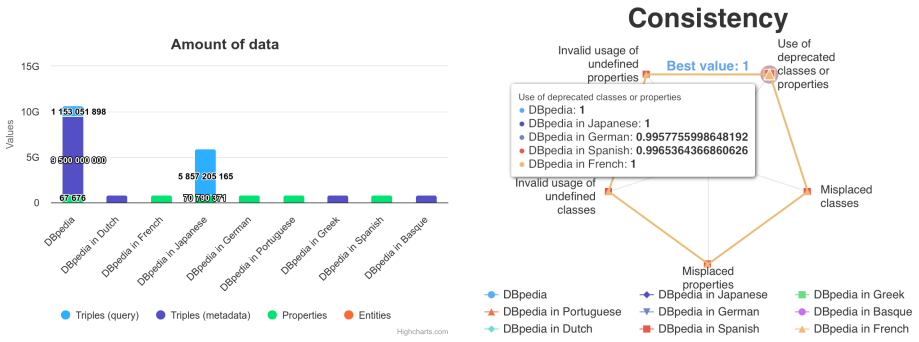
(a) Quality score table

KG name	2024-02-25	2024-02-18	2024-02-11	2024-02-04
DBpedia in Japanese	Online	Online	Online	Online
DBpedia in German	-	Online	Online	Online
DBpedia	Online	Online	Online	Online
DBpedia in Spanish	Online	Online	Online	Online
DBpedia in French	Online	Online	Online	Online
DBpedia in Portuguese	-	Online	Online	Online
DBpedia in Greek	-	-	-	-
DBpedia in Dutch	-	-	-	-
DBpedia in Basque	-	-	-	-

(b) SPARQL endpoint availability

Fig. 2: `KGHeartBeat` charts to compare linguistic DBpedia versions.

[10] Demonstration videos: `https://github.com/isislab-unisa/KGHeartbeat/tree/main/examples/videos`



(a) Amount of data via stacked charts (b) Consistency via radar charts

Fig. 3: (...continue) KGHeartBeat charts to compare linguistic DBpedia versions.

Acknowledgments - This work was partially supported by the SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

1. The SERICS project, https://serics.eu, [Online], Last access April 2024
2. Beek, W., Rietveld, L., Bazoobandi, H.R., Wielemaker, J., Schlobach, S.: LOD laundromat: A uniform way of publishing other people's dirty data. In: ISWC. pp. 213–228 (2014)
3. Debattista, J., Auer, S., Lange, C.: Luzzu-a methodology and framework for linked data quality assessment. Journal of Data and Information Quality **8**(1) (2016)
4. Esteves, D., Rula, A., Reddy, A.J., Lehmann, J.: Toward veracity assessment in RDF knowledge bases: An exploratory analysis. ACM Journal of Data Information Quality **9**(3), 16:1–16:26 (2018)
5. Käfer, T., Umbrich, J., Hogan, A., Polleres, A.: Towards a dynamic linked data observatory. Proc. of the LDOW Workshop at WWW (2012)
6. Keil, J.M.: ABECTO: An ABox evaluation and comparison tool for ontologies. In: The Semantic Web: ESWC Satellite Events. pp. 140–145 (2020)
7. Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R.: Databugger: a test-driven framework for debugging the web of data. In: 23rd International World Wide Web Conference, WWW. pp. 115–118 (2014)
8. Langer, A., Siegert, V., Göpfert, C., Gaedke, M.: SemQuire - assessing the data quality of linked open data sources based on DQV. In: Current Trends in Web Engineering - ICWE International Workshops. vol. 11153, pp. 163–175 (2018)
9. Vandenbussche, P.Y., Umbrich, J., Matteis, L., Hogan, A., Buil-Aranda, C.: SPARQLES: Monitoring public SPARQL endpoints. Semantic web **8**(6), 1049–1065 (2017)
10. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for linked data: A survey. Semantic Web **7**(1), 63–93 (2016)
11. Zhou, X., Zafarani, R.: A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Survey **53**(5) (2020)