paper_id	comment	analyst	sentiment
2801	This is the second revision of the article �Urban IoT Ontologies for Sharing and Electric Mobility�.	nan	nan
	The paper describes a modular suite of ontologies representing data gathered from Urban IoT devices used in urban mobility, finally producing a conceptual model to harmonize data exchanges between municipalities and service providers, with a specific focus on sharing and electric mobility domains. The paper also describes the methodology followed for the development of the ontology and contains a set of examples and references to additional materials to better understand installation, queries, and evaluation of the model.		
	In the previous revision I emphasized the clarity of the exposition, and the relevance of the covered topics. In fact, no major changes to the structure of the paper were required, since it was already well organized. As far as the writing is concerned, therefore, authors basically corrected the typos and missing references that had been highlighted in the review.		
	However, in my previous review, I underlined some shortcomings in the section of result evaluation dealing with Completeness. The authors have now added the file CompetencyQuestion_Completeness.xlsx to the repository, including numerous Competency Question examples on most of the classes of the ontology. This is indeed a useful tool to test the Completeness of the model, therefore my only advice is to replace the full file with a link to a google sheet in read-only mode. I suggest this because from git hub it is not possible to view an xlsx file directly, unless you install git hub desktop: you can only download the excel locally, which may not be the ideal solution for users and information security.		
	As for the long-term stable URL for resources, these are all organized in the git hub repository, introduced by a bilingual read_me. Moreover, the subfolders contain read_me description in English and/or properly commented code. The contents of each file are clear and consistent with the descriptions. I give a more than positive assessment to the material provided, and the same suggestion made for the CompetencyQuestion_Completeness.xlsx file applies to all the other xlsx files.		

This new paper also complies with all my other suggestions.

In conclusion, my final recommendation is to accept the paper. Indeed, such an example of integration of ontologies and the use of existing vocabularies meets the needs of the users it refers to and clearly describes the context of reference, resulting in a good quality and relevance model. Moreover, its modularity also allows for future extension.

Therefore, although its originality is not extraordinary, I appreciate not only the effort at the technical level, but also the detail and care in organizing heterogeneous data in the repository, and in explaining the problems and challenges faced and the results obtained. In fact, the work is complete and well documented.

2819	In this paper, authors introduce a novel approach which encompasses the aspect of interoperability to quality assessment of RDF data by identifying eleven interoperability dimensions by aligning a list of standardized dimensions regarding data quality to the �I� principles out of other FAIR data principles.	nan	nan

Authors also presented a tool by claiming its benefit for successfully identifying interoperability problems and could provide suggestions for resolving them.

It is a well written article and easy to read but following few points can be considered:

� There is no discussion around the validation and evaluation of the approach. It would be great to see the precision and recall of the defined approach especially in terms of complexity analysis and accuracy of the approach as the size of the data grows i.e larger datasets.

� As a proof of concept authors performed the quality assessment on a registry dataset about Addison�s disease and two synthetic datasets respectively about personal information and diagnosis from the European Joint Programme on Rare Disease (EJP RD). Addison's dataset was converted from a synthetic tabular dataset to an RDF dataset claiming the conversion scenario as typical/ useful

for quality assessment.

Typically, in most of the cases converting relational databases/ tabular databases to RDF requires a good amount of domain knowledge especially when it comes to medical/ biomedical data sources. Considering the inherent feedback mechanism from domain experts during the conversion process the quality assessment and the results may differ on a case to case basis.

How realistic would it be to comment on the quality dimension based on metrics defined in this paper at the generalised level i.e, on other RDF datasets. Would the same approach, dimensions and metrics work for any RDF resource?

Nevertheless, it would be great to see results in real world settings (on existing RDF data sources) irrespective of domain, complexity, size and coverage in order to consider the quality metrics defined in this paper for interoperability dimensions.

� Most of the data sources are dynamic in nature these days. Does any of the quality metrics defined for interoperability dimension consider the assessment for such scenarios?

� Twenty-nine failure types were defined in the Interoperability Failure Case (IFC) vocabulary which seems comprehensive but it is not explicitly clear if IFC is one of the contributions by authors or the re-use of existing vocab named �IFC�.

� Would be good to see the false positive results while considering the failure matches based on the defined technique. Any analysis based on computed results would be great to see.

� Some of the possible future directions are highlighted in the last paragraph of section 5 and would be good to present them as separate sections/ subsections for better visibility.

� Clear statements regarding the limitation for the approach, data used and the developed tool should also be highlighted.

Despite the consideration made by the authors in the cover letter,

2833 this paper must be reviewed as "Tools and Systems Report". The paper is far from a full research paper and does not add relevant scientific contributions to the Semantic Web field; instead, it uses existing tools to create an event-based knowledge graph.

nan nan

The topic addressed in the paper is very relevant, timely and well-introduced and motivated, inspiring its reading. On the other hand, the related work section is long, does not explain the paper's contributions in relation to the state-of-the-art and is difficult to follow.

The authors mention "first step for building a knowledge graph was to decide the source of the documents, since there are important differences among jurisdictions, even when they share the language". A suggestion to improve readability is to provide an example. I can only guess what the differences are.

The authors also mention "From the analysis performed in the EventsMatter corpus, we can confirm the importance of the sections in identifying which events are relevant and which are not". It would be better if the authors would let the readers 'confirm' by providing in-depth analysis and supporting data. The figures presented do not seem enough to confirm the importance of the sections rather than the frequency of facts.

The authors also mention that the "Structure Extractor is currently able to handle the structure of the ECHR and ECJ documents, but in such a way that a new document type can be easily added". Looking at the code and the lack of documentation/comments, it does not seem trivial to change a single line in the code provided. Paths are hard-coded, filenames are hard-coded and no design pattern was found to handle the extensibility claimed in this paragraph. An example is the method "parseAndTag". In general, section 3 is confusing and terms are mentioned without a proper introduction. For example, the authors say the EventsMatter corpus and only in a subsection it is 'properly' explained (same for the FrameNet). Section 3 requires restructuring. As mentioned before, a running example would help the authors to explain the process.

The presented ontology is also not well-explained. For example, the class "ComposedTemporalExpression" can be represented in other

ontologies. Simplification can lead to other issues. Again, examples would allow readers to see the 'importance' of such class.

The remaining sections suffer from similar issues.

Another main concern about the paper is the evaluation of the tool. The authors did not provide any evaluation/experiments�a critical point for a tool.

As previously mentioned, the code is available but not easy to run, lacks documentation and coding standards are non-existent (https://www.oracle.com/java/technologies/javase/codeconventions-contents.html).

2852	1. Originality:	nan	nan

This paper introduces a neural network model where attention mechanisms are exploited to learn common object properties from the OWL2 ontology standard.

Similar to what some of the co-authors proposed for the IterE model, this paper exploits axiom properties throughout its learning, but instead of learning rules guided by those properties, it uses those rules to implement an attention mechanism affecting the model embeddings.

In conclusion, it exploits recent deep learning techniques, trough attention mechanism, along with Ontology formalisms, which represents an original way of mixing deep learning and symbolic AI.

2. Significance of the results:

This paper provides tangible proofs about the value of ontologies' axiom information, in the learning process of knowledge graph embeddings.

Information on test results and the method used to produce them are well presented.

Results on the WN18RR with Hits@1 metric, being a very difficult task, are very good, and show potential advantage of this method for tasks where precision is needed in noisy (real life) settings.

The significance of those results is satisfactory but somewhat modest considering additional tests and evaluation would be needed to fully assess the level of improvement brought by this method on knowledge graph embeddings and its applicability to different type

of datasets.

Nevertheless, it truly confirms the validity of the approach and applicability to certain use cases.

3. Quality of writing:

The quality of the writing is very basic, many sentences are missing words or using the wrong word, verb tense, etc., which makes reading unpleasant but does not affect the technical comprehension.

4. Paper Code/Resources:

The github reference was created but is empty (at least the public branch), which makes it impossible to review the solution code quality, reproduce results and confirm whether artifacts will be available.

2866   The manuscript was submitted as 'Data Description'. It describes a linked dataset of SPARQL query logs (LSQ), version 2.0. LSQ 2.0 features logs from 27 SPARQL endpoints and describes 43.9 million SPARQL query executions on these endpoints. The paper is an extension to a paper "LSQ: The Linked SPARQL Queries Dataset" published on ISWC 2015. The submited extended version has the same structure as the previous paper. Both papers consider the same set of use cases for the LSQ dataset and both introduce the same data model for representing query logs. The reviewed paper describes significant extensions compared to the first version from   nan   nan
2015. This includes a significant extension with more datasets, queries and query executions, and more detailed statistics. Also a broad usage of the dataset is presented in a structured form which transparently shows various recent research works which used LSQ for the evaluation. The paper shows only the usage of LSQ 1.0. However, this clearly proves the usefulness of the dataset. The new version as a significant extension to the first version will undoubtedly find a broad usage as a pool of queries for various benchmarks.

In summary, even though the paper does not bring anything ground breaking compared to the paper from 2015, it shall be published in my opinion. It is necessary to inform the research community about the new version of the dataset. The paper contains all necessary information the readers need to start using the dataset.

I have two minor issues related to the information provided about the dataset in the paper.

I1) Metrics and statistics on external and internal connectivity of the dataset: The SPARQL endpoints are described in a LSQ specific way. However, existing descriptions of SPARQL endpoints expressed as VOID resources could be linked instead.

I2) Use of established vocabularies: This is related to the previous issue. On page 4, Query instance, the authors describe that the originating endpoint of a query is associated with the dataset using a LSQ specific property lsqv:endpoint. The reason is that void:endpoint or sd:endpoint could not be used because they have different domains. It is a strange reason. For example, a LSQ query can be associated with a void:Dataset using a LSQ specific property. Then it would be possible to use void:endpoint (https://www.w3.org/TR/void/#sparql). I understand that it is not possible now to change the vocabulary. However, the rationale behind this design decision should be justified better.

The paper demonstrates the necessary level of the maturity of the dataset, including its quality, stability, and usefulness proved by the usage of the dataset by many researchers as a benchmark for evaluating their results. The description of the dataset is at the necessary level of detail. I have two minor issues related to the clarity of the description.

I3) Page 4, query instance: The property lsqv:endpoint is described in the text with lsqv:Query as a domain. However, Fig. 1 shows that the domain of lsqv:endpoint is lsqv:RemoteExec. The textual description of the LSQV vocabulary is inconsistent with the figure.

I4) Page 4, static features: It is described that static features are defined for a query, independently of the dataset over which the query is evaluated. However, the paragraph above describes that queries are considered only in the relation to an endpoint. In other words, an instance of lsqv:Query is always a pair of SPARQL expression and endpoint. LSQV does not allow to describe features of queries independently of the endpoints. Therefore, the paragraphs "Query instance" and "Static features" on page 4 are in a contradiction.

Regarding the LSQ data file

(A) It is well organized and available for download as well as via a SPARQL endpoint through an access page http://lsq.aksw.org/. The access page contains necessary information about the access to the dataset.

(B) The provided resources are complete for replication of experiments. The access page provides also tools for the experiments and their documentation.

(C) LSQ is not available through a public repository such as GitHub, Figshare or Zenodo. The authors use their own web server which seems appropriate for long-term discoverability. LSQ does not have a DOI.

(D) The provided data artifacts are complete. The LSQ dump is structured to files for individual SPARQL endpoints. The content for each endpoint seems complete.

| 2885 | "(1) Quality, importance, and impact of the described tool or system (convincing evidence must be provided). " | nan | nan |

This paper presents TermItUp, a generic architecture integrating multiple state of the art tools with the purpose of providing a one-stop-shop for all terminology extraction needs.

The tool has been developed following FAIR and open science principles, using standard LLOD and LOD formats, guided by a set of requirements based on observations in the state of the art,

but also discussions with terminology experts.

The tool will be an extremely useful to the community as the systematic integration of its different components for each single project would be incredibly time-consuming and ad-hoc.

There are high impact uses of the tool in H2020 and other collaborative projects, showcasing the potential of the tool.

"(2) Clarity, illustration, and readability of the describing paper, which shall convey to the reader both the capabilities and the limitations of the tool. Please also assess the data file provided by the authors under �Long-term stable URL for resources�."

The paper is well written and organized, and clearly tell the story of TermItUp in terms of its capabilities and shortcomings.

There are small language corrections to be made for the camera ready version, highlighted at the end of this review.

The review of literature is complete, although I would have preferred to see at least a mention of SOTA efforts for terminology-extraction around TermEval2020/ACTER and an explanations as to why such systems although theoretically very accurate would be very difficult to integrate in a production system.

There's a very interesting multilingual extraction system in the 2021 finding of the ACL https://aclanthology.org/2021.findings-acl.316.pdf

I mention this because it's interesting, but the positioning of the paper doesn't necessarily require to go into this particular literature.

Perhaps the mention in the paper that there are ongoing efforts for multilingual terminology extraction actually refers to this.

Regarding the perspectives, I would love to see the future integration of a knowledge-graph aware association rule mining approach in addition to the extraction of hierarchical relations.

I am not asking to mention this in the paper, just an interesting thought.

"In particular, assess (A) whether the data file is well organized and in particular contains a README file which makes it easy for you to assess the data,

(B) whether the provided resources appear to be complete for replication of experiments, and if not, why,

(C) whether the chosen repository, if it is not GitHub, Figshare or Zenodo, is appropriate for long-term repository discoverability, and

(4) whether the provided data artifacts are complete. Please refer to the reviewer instructions and the FAQ for further information."

The github repository is easily accessible and significant information is given in the README or on the main website of the tool.

The overall explanations are clear and the documentation of the API provided through swagger is very helpful.

The main improvement directions would be 1/adding a more technical documentation that explains how to deploy the service 2/adding some docstring documentation for developers

3/potentially refactor the code base if there is a subsequent increase in complexity, I would personally favour a generic class structure with polymorphic genericity rather than reflexively loading code modules that all include the same functions, even though the latter tends to create less overhead.

As the tool is still a prototype under active development, those are not significant issues with regard to the publication of the paper.

I recommend acceptance of the paper, the corrections are mostly cosmetic for the camera ready.

Detailed corrections

Page 2 line 12 left: the bit of the sentence about DBPedia is confusing, I don't understand what it means.

P2 l19 left: is to find -> is finding

P2 l25: different backgrounds and expertise levels to face language and related needs [...]

P2 l30 right: discussions that have arisen

P2 l37 right: This section covers

P2 l38 right: different processes mobilized in our system

P3 l42-43 right: Combining Wikipedia and other resources, BebelNet constitutes an multilingual [...]

P4 l9 left: domains, with half being closely related to [...]

P4 l10 left: Several scientific works are devoted

P4 l12 left: a SPARQL

P4 l28 left: can be of great help

P4 l6 right: corpus -> corpora

P4 l36 right: These can correspond to different [...]

P4 l43-45 right: The meaning of a unit is to be discovered in text and constructed through relations to other terminological units.

P5 l22 right: can significantly contribute to improving performance [...]

P5 l22 right: to developing

P5 l45 right: this translates to a necessity [...]

There are aditionnal small corrections like these to be made, can transmit feedback to authors later as time permits, preferable to delaying the review submission.

2898

The authors have exhaustively addressed all previous comments, and I was very happy of the improvements.   nan   nan

The paper is much clearer and easier to follow.

The presentation of the two methods used to align discourse segments and relations across the parallel corpora is clear and sound - description of Method II has improved a lot.

Editorial comments:

- move Table 2 to be closer to where it is referred to in the body of the document

- move Table 3 to be closer to where it is referred to in the body of the document

- pag 8 line 14: incriminating --> increasing?

- pag 9 lines 15/18: check for consistency the use of italics for the connectives in the body of the text

- pg 10 line 20 and following: when presenting the sources of errors, I would use the \paragraph{} command in LaTeX rather than starting the paragraphs with a sentence in italics

- Table 9 and Table 10: I would anticipate and discuss Table 10 first - move it to where Table 9 now is and then present and discuss Table 9

| 2904 | The paper presents the work to properly publish data under the principles of linked (open) data). To do so, the authors have promoted to the linked data initiative a dataset coming from Nova Scotia. | nan | nan |

The work is interesting and applicable to many domains in which sometimes data is published for the shake of publishing. However, there are some things to improve:

-In the abstract, it should be necessary to be more informative and quantify some adjectives like �most of datasets� to see some context and magnitude of the problem/context.

-Include also more technical details about the results: ontologies/semantics rules and provide some quality assessment if possible.

-In the introduction, some existing issues must be justified: "The datasets act as isolated pools of information that cannot be queried or linked." This implies to categorize and identify the current dimensions of the problem to be addressed:

--Problem of reusing data, Which are the principles? For instance, alignment to Open Science/Data principles, etc.

Then, establish the specific technical issues that are preventing a proper data federated environment: data modelling: linked data patterns? data integration, data quality, data querying mechanisms, lack of APIs, use of standards and existing vocabularies, etc. to finally assess if there are interoperability issues: interoperability issues: communication protocols, syntax and semantics

-In the state of the art/background section, there are multiple works about linked data lifecycles that must be mentioned.

-In the methodology section, it would be nice to see some description of the datasets (type of data, issues in data: consistency, naming, etc., need of logics, etc.) to finally end with the decision on the vocabularies and data model including a description of what is being reused (with more details of the process) and what is new.

-Regarding the semantic rules, does it make sense to directly use SPARQL to produce new facts? In terms of consistency, it should be

nice to see the reconciliation process between entities (if it was necessary).

-Authors also show some SPARQL queries. It should be necessary to link the potential of these queries to the initial needs.

-The quality checking of the dataset seems a bit simple. It would be also nice to see how to consume that information (if it is publicly available).

Finally, authors properly comment the main conclusions and envision some future work.

| 2911 | (1) Originality: | nan | nan |

The paper describes the Intelligent Energy Systems Ontology (IESO), which aims at supporting an interoperable co-simulation of the power system including market mechanisms. The authors proposed a so-called multi agend system (MAS) society build of different MAS and ontology-based communication. With the ontology the data models of these different MAS are defined and validation and unit conversion are supported. The approach of a ontology-based co-simulation in energy systems context is novel and promising. The used MAS and their ontologies, were already desribed in earlier papers of the authors and are integrated into the IESO.

(2) Significance of the results:

The authors give an overview over already available ontologies in the domains of interest and describe how they make use of them for their approach to support interoperability with other tools. The modules of the IESO are described detailed, mentioning the dependencies between the modules and to other ontologies. In the explanation the authors mention, that "owl:sameAs" is used for referencing other ontologies, but I did not find this in the online available version of IESO and the related ontologies.

A case study shows the capabilites for the simulation of a distribution grid with some households and market mechanisms. The data of every step of this case study and the used SPARQL queries are provided at Zenodo, which makes it very transparent. Two of the used MAS are available as web services, but the other used tools seem to be not openly available. Thus, the results can be retraced,

but as the tools are not available and there is no instruction how to execute the tools, it is not possible to reproduce the case study and some questions stay open. How are the provided queries and data between the different tools exchanged? Manually or by script/TOOCC? How complicated would it be to change components of the simulation?

Additionally, an evaluation of the performance of the approach would be interesting. The case study is focusing on a quite small grid with the simulation of only one time step. Of cause, this is sufficient and suitable for demonstrating the process and ontology, but for realistic simulation of the power system large systems have to be considered. Thus, at least some considerations about the scalability of the approach for larger grids, more agents, or simulation of larger time intervals would be helpful to value the relevance of the approach.

(3) Quality of writing:

Overall the paper is written well, the approach is described in a well understandable way and figures are used to visualize it. Many figures have a quite low resolution and should be replaced by vector graphics or higher resolution source files.

Some small remarks:

- line 41-44: "... such as: ..., to name a few" doubled

- line 19: Oxford comma missing after "operation"? "simulation" instead of "simulations"?

- line 258: Is "hardening" the right word here? Maybe better "hindering"?

- figure 1: There are used two different nuances of orange and different fonts (e.g., boxes "Decision Support" and "Historic Data"). If this is deliberately, it should be explained. Otherwise it should be consistent.

- figure 1: Some parts of the figure are barely readable, even when zooming in. A vector graphic should be used or at least a high resolution source file.

Open Science Data:

(1) Organization of the file:

In the provided data file a README file is missing. <mark>The description of the case study in the paper contains references to the according files in the data, which allows to follow well.</mark>

(2) If applicable, whether the provided resources appear to be complete for replication of experiments and if not, why:

<mark>The provided data seems to be complete to allow reproducing the case study.</mark> The Power Flow Service and Electricity Market Service are also available as Web Serives, but the additional used tools (TOOCC, MASGriP, AiD-EM, SSC) seems to be not available and there is no instruction how to use the different tools to execute the case study. <mark>Thus, from my point of view, it is not possible to reproduce the case study.</mark>

(3) Assess the appropriateness of the chosen repository if it is not GitHub, Figshare, or Zenodo:

The IESO and all of it's modules are hosted on a homepage and have integrated the version number in the URL. Due to not using a repository like GitHub, FigShare or Zenodo, it seems that version management and repository-discoverability is not supported and no DOI is used.

(4) If applicable, whether created data artifacts appear to be fully included in the file:

<mark>According to the description of the case study in the paper, the data artifacts seems to be complete for all the described steps of the process.</mark>

2914 | This manuscript showcases a Stream Reasoning application in the healthcare context. <mark>It is well written and appropriately organized. The goal of the research is clear and relevant.</mark> <mark>The originality is low.</mark> The same authors published a similar paper in [1] on September 2021. <mark>The significance of the results is also low.</mark> The paper presents a use case. <mark>It demonstrates that it is possible to solve the problem using Stream Reasoning, but it does not prove that it makes sense.</mark> <mark>I would expect to see a comparative study that shows advantages and disadvantages compared to a solution purely based on a stream processor engine (e.g., esper, ksqldb, flink, or Spark Structured Streaming).</mark> nan nan

The data file provided by the authors under �Long-term stable URL for resources� is well organized and contains a README file. The provided resources are published in their institution GitLab and appear to be complete for replication of experiments. I did not import the project using an IDE, but the artifacts appear complete.

As a side comment, I invite the authors to stop using c-sparql and use rsp4j [2,3] instead.

[1] Mathieu Bourgais, Franco Giustozzi, Laurent Vercouter: Detecting Situations with Stream Reasoning on Health Data Obtained with IoT. KES 2021: 507-516

[2] Riccardo Tommasini, Pieter Bonte, Femke Ongenae, Emanuele Della Valle: RSP4J: An API for RDF Stream Processing. ESWC 2021: 565-581

[3] https://github.com/streamreasoning/rsp4j

| 2915 | The paper can be accepted. | nan | nan |

I have however some recommendations for future work on the subject.

1) Obviously it is unlikely that the first version of an ontology is perfect. It will be necessary to refine what at present seems a very complex structure, with some possible redundancy in the class & property definitions. One example: why there is a new class FormationProcess, subclass of A4 Stratigraphic Genesis? Was the latter insufficient? Was this new class strictly necessary and a solution with "P2 has type" not suitable? Thus I would consider the paper a first attempt, requiring further work to be implemented but still worth publishing. In the meantime, implement Occam's razor!

2) The repository part is still very preliminary. This is acceptable at the present level of development, where the focus is on the theoretical aspects of the ontology and not on its implementation, but it will be necessary to go beyond Google drive (as authors also state). I would not consider this a secondary aspect in the future, nor an easily solvable one.

3) Same comment for the use of Omeka-S. Is it the appropriate tool with all the required functionalities? I am not sure it is powerful enough. The choice of the package should go hand-in-hand with the repository solution.

However, such data management aspects (points 2 & 3) can be addressed in future work and it is useful to publish the present one as is, not least to allow a discussion on the matter and compare different solutions.

| 2918 | * Summary: The article describes an approach to convert vector geographic features extracted from multiple historical maps into contextualized Spatio-temporal KGs. The resulting graphs can be easily queried (GeoSPARQL) to understand the changes in different regions over time. The approach and its evaluation focus on linear and polygonal geographic features. | nan | nan |

* Overall Evaluation (ranging from 0-100):

[Criterion 1]

[Q]+ Quality: 95

[R]+ Importance/Relevance: 85

[I]+ Impact: 90

[N]+ Novelty: 80

[Criterion 2]

[W]+ Clarity, illustration, and readability: 95

[Criterion 3]

[S]+ Stability: 100

[U]+ Usefulness: 90

[Perspective]

[P]+ Impression score: 90

* Dimensions for research contributions (ranging from 0-100):

(1) Originality (QRN): 87

(2) Significance of the results (ISU): 93

(3) Quality of writing (QWP): 93

* Overall Impression (1,2,3): 91

* Suggested Decision: [Accept]

* General comments:

The work is solid. The paper is easy to follow and understand.

There is a novelty in the proposed approach.

* Major points for improvements:

{

all have been addressed

}

* About the data files and related software artifacts: (�Long-term stable URL for resources�)

(1) "data files are well organized and contain a README file which makes it easy to assess the data": [YES. CORRECTED]

(2) "the provided resources appear to be complete for replication of experiments": [YES. COMPLETE]

(3) "the chosen repository is appropriate for long-term repository discoverability": [YES. GitHub]

(4) "the provided data artifacts are complete": [YES]

* Minor corrections:

[A06] Indeed, this (unrotated) box is the required input to the reverse-geocoding service and poses a limitation for us. We have clarified it and described how the calculation is done (section 2.4, paragraph 5).

[R06] In the manuscript, you didn't mention that is a "limitation" of the reverse-geocoding service. Perhaps, it would be better to mention this part in section 5, paragraph 5, where you described "several limitations".

| 2924 | The paper tackles the problem of knowledge graph (KG) construction. It proposes planning and execution techniques to speed up KG construction pipelines specified using mapping rules in [R2]RML. The proposed method relies on the partition of mapping rules so that the evaluation of the groups in the partition, reduces the duplicated generation of RDF triples and maximizes the parallel execution of the mapping rules. The proposed techniques are implemented in MorphKGC, an RML-compliant engine; the behavior of MorphKGC is assessed in three existing benchmarks. | nan | nan |

The reported results suggest that the proposed methods can accelerate the execution of mapping rules as the ones composing the studied benchmarks.

Overall, the paper is relatively well-written and presents an efficient solution to a relevant data management problem. The experimental results provide evidence of the benefits that planning the execution of the mapping rules brings to the process of KG construction. Moreover, the outcomes of the empirical evaluation put in perspective the improvements achieved by the proposed planning techniques. However, the current version of this work suffers from several issues that considerably reduce its value. First, the problem is not formally defined, the definitions are not well-formulated and several relevant concepts are idle defined. Second, the complexity and correctness of the proposed algorithms, and the conditions that prerequisite their performance and soundness are not even mentioned. Third, the experimental study is conducted over a limited set of test cases, preventing, thus, the reproducibility of the reported outcomes. Finally, the state of the art is superficially analyzed, and the proposed techniques are not positioned with respect to similar data management techniques.

All these issues prevent from a positive evaluation of the current version of the work. The recommendation is for a major revision addressing all the following comments.

Mathematical Pitfalls:

� Definition of equivalent triples maps. The conditions to be met to decide when two triples maps are equivalent are not formally stated. Note that two triples maps can produce equivalent results even if they are defined over two different logical sources. However, the informal definition presented on page 3, suggests that both triples maps should be defined over any given data source. Moreover, the property of �two equivalent term maps with different positions are not equal�. The differences between equivalent and equal must be differentiated. What is the complexity of the problem of deciding if two triple maps are equivalent?

� �Because of RDF set semantics�. An RDF document is formalized as a graph, please, clarify and reference to which RDF set semantics the authors refer to.

� It is never defined when an RDF triple is generated from the evaluation of a triples map over a logical data source. Please, formally state the results of a triple map evaluation.

� Definitions of position(.), type(.), value(.), and literaltype(.) incorrectly assume in the domain a single element T. Contrary, the domain in all these cases must be a set of term maps.

� Self-joins are used without any definition in the context of mapping rules.

� Definition 1. The concepts [R2]RML document, equivalent normalized [R2]RML document, and [R2]RML documents without self-joins are not used without any previous formal definition. Consequently, the description of canonical [R2]RML document is mathematically incorrect.

� Definition 2. A partition of a set X is a set of subsets of X, and the relationship of the parts G_i of P is not defined. Also, please, note the time complexity of enumerating all the possible subsets of a set X.

� Definition 3. A template should be defined in terms of the different structures that it may take, and then a prefix can be defined.

� Definition 4. An invariant is wrongly defined. T is a term map, while an invariant I is a string, and making I equal to T is a type mismatch. Also, the �if� conditions only state sufficient prerequisites. Please, provide sufficient and necessary conditions of the values of an invariant I.

� Definition 5. Term maps are considered sets, even though a term map is never presented in this way

� Definition 6. The definition of Disjoint mapping rules relies on the definition that a triple set is generated from a mapping rule which has never been defined.

� Definition 7. It presents a property instead of the definition of a Disjoint mapping Groups. Also, it relies on definition 6 (which is incorrect) and assumes that a triple map is a set.

� Definition 8. It defines a Maximal Mapping Partition of an [R2]RML document as the one with the largest number of mapping groups. This is exactly the one where each group is a singleton set unless another missing condition needs to be satisfied. What is the

complexity of finding a Maximal Mapping Partition?

Proposed Algorithms:

� Algorithm 1 simply replaces an object reference for the template definition of the corresponding parent triples map. Algorithm 2 and 3 use functions which are not defined.

� Heuristics implemented by Algorithm 3 are not defined and because the problem solved by Algorithm 3 is not defined, it is impossible to demonstrate its correctness.

� The complexity of Algorithm 3 requires to be demonstrated.

Empirical Evaluation

� The empirical evaluation does not consider engines that also implement similar planning techniques, e.g., SDM-RDFizer 4.0 [1].

� The meaning of Table 1 is not clear, and the reported results do not have any connection with the rest of the outcomes presented in this section. Please report the results if the experimental study was also conducted over all these datasets. Otherwise, eliminate this table. It is misleading for the reader.

� The absolute values of figures 4 and 5 should be reported.

� Detailed discuss of the conditions to be met by a data integration system to benefit from the proposed techniques. For example, what happens when the joins between triples maps are not self-joins, and the selectivity change? The authors claim that they have discussed the parameters that impact the execution of the KG construction process in previous work. However, only a few of them are considered in this study, reducing, thus, reproducibility of the results in more general testbeds.

Related Work

� This section simply describes tools instead of analyzing the data management techniques proposed by each of the existing approaches. Please, provide a deeper analysis of the problems solved by the approaches reported in the literature, their pros and cons of the proposed techniques, and position your techniques with respect to them.

Minor comments

� The benchmark Genomics � TIB is mentioned but then, the

COSMIC testbed is used. Are they both the same? Do the authors refer to the benchmark available at [2] or at any other benchmark?

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

[2] https://figshare.com/articles/dataset/SDM-Genomic-Datasets/14838342/1

| 2926 | The goal of this article is the development of an ontology NagO. NagO is a domain ontology focused on terms regarding the United Kingdom and its external territories and their link to the Nagoya Protocol policy framework, processes and people involved. | nan | nan |

The abstract clearly and concisely presents the purpose of the article. The main idea for developing the Nagoya Ontology (NagO) is to semantically model the complex policy framework around the Nagoya Protocol and to unveil the legal relationships between sovereign states and their external territories, illustrating the United Kingdom as a study case.

I don�t have any comments here.

The introduction very clearly presents the reasons why the authors have decided to develop the Nagoya Ontology. Obtaining heterogeneous information about the sovereignty of different external territories of countries and not free access to reliable information on the status of these territories under the Nagoya Protocol gives the authors reason to emphasize the need to develop Nagoya Ontology.

I don�t have any suggestions here.

Quality and relevance of the described ontology.

I think the ontology is qualitative and relevant.

1. The information in NagO is carefully selected and it follows the Basic Formal Ontologyconcept and the guidelines under the OBO Foundry. For the sovereignty branch of NagO, the Island Rights Initiative provides the majority of information by displaying the constitutional links and governmental relationships of every external territory of the United Kingdom. The approach to collecting and presenting the information contained in NagO is described systematically.

2.The ontologies, which provide terminology for NagO, include

ENVO[16], IAO[17], BFO[8] and RO[18], and they were automatically imported using the ontology development kit [19]. The Dublin Core Metadata Initiative terms [20] and NCBITaxon [21] have been added with a manual import in Proteg�.

3.Specifically looking at the term definitions gave the authors an insight of how these terms are related to each other. The basis for this step was reading about administrational and political processes and unveiling the distinct people involved, the roles they have and what input is needed for certain processes. The ABS Clearing House offers information on country profiles and documents in relation to the Nagoya protocol.

4.Where public data information was not found, personal inquiries were made to the government departments. The OBO Foundry delivers definitions and relations for the biodiversity and territory branches of NagO (ENVO, SDGIO, GO, NCBITaxon and GEO, GAZ respectively).

5.In Figure 1, the authors represent an overview of the NagO domain scope intersections with an associated topic and possible connections to existing ontologies. In the �Key Futures� Section they present and describe the connection of NagO with other existing ontologies in different domains connected with the Nagoya Protocol and the identify niches where semantic development is needed.

6.The authors define some semantic patterns in the ontology. An example can be seen in Figure 3 - geographic territories are related to people and processes. There are three more semantic patterns defined in the ontology: Territory instances and the execution of a certain government type (e.g. self-governance); UK external territories and their connection to the United Kingdom; Ministerial departments and their connection to the head of state. All these semantic patterns can be used in other domains.

7.With the example for use of the ontology, the authors demonstrate that NagO can be used to query constitutional links of geographic regions. They also claim that it can be used to find role capacities of the people involved and relevant processes and instances of the Nagoya Protocol.

8.In the last section of the article, different opportunities are discussed to extend NagO in different domains.

9.NagO is free and openly accessible in English on
�https://github.com/hseifert/NagO�.

Illustration, clarity and readability of the describing paper, which
shall convey to the reader the key aspects of the described ontology.

I think the paper is well written and all ideas are adequately
presented with figures and tables. The developed ontology is
properly described and the content of the article is clear and
readable. All Figures and tables are explicit and comprehensible.

All files � ReadMe, nago.owl, nago-full.owl and the others are
freely available at https://github.com/hseifert/nago/tree/develop

Review comments:

1.My major comment is about the �Background� section. In this
section, the authors again present the reason to create the NagO
ontology and its connection with other domains. I think it will be
interesting to the readers if there are some similar ontologies about
other protocols (conventions) to be referred here.

2.The methodology for creating the ontology and gathering
knowledge is well described. Various domains related to the Nagoya
protocol are presented. These domains are included in the NagO-
related ontology. There is a need to describe the connection to these
related ontologies. Which classes are really re-used and how do the
different ontological commitments relate to each other?

3.�Both figures 2 and 3 include shortcut relations used for this
illustration� � I didn�t see the shortcut relation in Figure 3.

4.I think the authors can extend the usage scenario, for example with
finding role capacities of people involved. This will more fully
present the potential of the NagO ontology.

The paper describes the authors' effort to define a modular ontology
for common sense analysis. Their ultimate goal is to increase the
density of the final knowledge graph by structuring and lowering the
complexity. The have utilized the relation in WorldTree and created
structured relations over those. Their ontology contains relations
such as: structured relations, Verbal relations, Taxonomic relations,
Affective relations, and grammatical exceptional relations. The
paper carefully defines the relation type included in each category.

2949   nan   nan

Afterward they have described the annotation procedure to be combination of manual and automatic procedures and provided comparison of statistics of their KG with other common-sense knowledge-bases. Finally for comparing the quality of the ontology and collected KG, the paper describes the experiments over the task of QA and an ablation study to show the quality of every category of relations.

In terms of Quality and relevance of the described ontology, the paper provides a very structured ontology which helped in improving the results of the QA based task, yet there can be tested on more sub-tasks.

The paper is clear and readable but as a suggestion an outline can be added to the content in the introduction sections.

The authors have shared a GitHub repository, containing their final knowledge-base, yet it is a little confusing and hard to navigate, it might be better to add README file. In addition the GitHub link does not include their scripts used for annotation. It might be good if the authors make those also accessible for other researchers to use.

The strengths and weaknesses of the article are as follows:

Strengths:

+ The authors gave a better structure to the existing ontologies in order to create a dense knowledge graph which ultimately will result in better inference.

+ Using the described ontology the authors were able to use a combination of manual and automatic annotation scheme which reduces the manual effort needed to create the knowledge-base.

+ The evalution presented in the paper shows a proof that utilizing more structured graphs (even with less nodes) can result in improvement of the model's performance in the QA task.

Weaknesses:

- There are not many qualitative samples evolutions provided by the paper.

- For the evolution, the comparison is drawn over the task of QA using QA-GNN, but it might be better to have the evolution over

couple of other tasks to establish the performance of the new ontology and the KG created via that.

Questions for the Authors:

* In the section, 3.3.2, it is mentioned that some concepts (e.g. distance and far) are kinda similar, is there a threshold for determining similarity?

* In the sentence : "bakers give bread to customers." The authors mentioned that the given triple will be (�bakers give bread�, beneficiary, �customer�), but does it also provide other artifact relations such as ('bakers', 'own', 'bread')?

* In section 4.3, the authors performed post processing in order to find and prune the errors. In what are the percentages of occurrences of those errors.

* For section 4.3, are there any other type of the errors that remained within the annotated samples that are not caught, it might be interesting to see more qualitative samples of the final datapoints.

* In section 4.3, page 14, line 6, there is a mention of "In others this was genuine (�open container� is both a possible verb and noun phrase)." but it is not clear what is the authors' approach for these cases.

* Table 10 shows that the results of KG analysis without taxonomic relations is even higher than the model with those. If possible can the authors provide the samples that are marked correctly without taxonomic relation and not with them?

* The authors have started with the WorldTree relations, Is this annotation approach applicable to conceptNet?

Minor Comments:

* It might make the paper more readable if the authors can add the outline of the paper of what to be expected in each section and the overflow of the paper.

* In page 7, line 18: base classes are combined in --> to be combined

* In page 8 line 31 there are 2 "cause" words.

2964 In this paper the authors describe the ExaMode ontology and some of its applications. ExaMode includes a few hundred terms for disease, diagnosis, patients, tests, test results, procedures, etc. focused on histopathology, and specifically on four diseases: colon cancer, lung cancer, uterine cervix cancer, and celiac disease. The system has been used to automatically extract annotations from case reports in Italian and Dutch, and the results can be visualized as a graph. nan nan

This work is valuable and difficult. The authors are to be commended for building a working system from many disparate parts, and for making good use of existing ontologies to improve interoperability of their systems and data. They provide a good ontology description in this paper, detailed ontology documentation at http://examode.dei.unipd.it/ontology/index.html and open code https://github.com/ExaNLP/sket/. I encourage the authors to continue this work.

Specific review requirements:

(1) Quality and relevance of the described ontology: I believe ExaMode is relevant, and suitable for purpose, although I have concerns about quality discussed below.

(2) Clarity of the paper: Good overall, with specific points below.

The attached Zenodo archive includes `examode.owl` and conversion to several other formats. (A) iI includes a description but not a README per se. (B) The files are complete ontology files. (C) Zenodo is appropriate for archiving. (D) No other data artifacts are provided.

General concerns

I do, however, have a number of concerns about how the ontology is built, which I hope will lead to clarifications and improvements of the ontology itself. These criticisms may also help with revisions of the paper.

This project seems to be squarely aimed at annotation, and I am persuaded that it is suitable for that purpose. When I look at the details of the OWL file, I see several problems. While it is good to reuse existing terms, ExaMode includes quite a mixture of terms

from different source ontologies with different modelling strategies. For example, UBERON logically defines 'endometrium' as equivalent to "mucosa and part of some uterus", but ExaMode asserts that 'endometrium' is `owl:partOf` NCIt's 'mucosa' term. The OWL specification does not include `owl:partOf`, but more importantly this should be a subclass relation, and I do not understand why NCIt's 'mucosa' term was used instead of UBERON's 'mucosa' term. Term reuse aids with interoperability, but inconsistent term reuse undermines that goal.

It was not clear to me from the paper, but looking at the OWL file in Protege I was quite surprised that UBERON and MONDO classes had been "demoted" to owl:Individuals. This was not always done consistently, so MONDO:0002271 'Colon Adenocarcinoma' is an owl:Class, but MONDO:0002032 'colon carcinoma' is an owl:Individual. It suggests that there were two 'Resection' procedures, but I suspect there was only one. For the outcome on the left we have 'Mild Colon Dysplasia' as the `rdf:type`, but on the right the same term is used as the object of a `exa:hasDysplasia` predicate, and 'Moderate Colon Dysplasia' is also specified.

More generally and more subjectively, ExaMode's four "semantic areas" seem to each have quite different modelling approaches, which become apparent when looking at the OWL file as a whole. For example, an 'Onset' is not a subclass of 'Annotation', although a record may be annotated with information about an onset.

Although the authors do make good reuse of many existing ontology terms, I believe that there are good candidate terms in OBO and elsewhere for many of the terms that they do create in ExaMode. I would encourage the authors to expand the scope of their collaboration, and grow ExaMode toward closer integration with the larger open ontology community.

Specific points about the manuscript

- Abstract and elsewhere: "four largely diffused and studied histopathology diseases"; "diffused" seems a strange word choice to me (as a native English speaker)

- Page 2 line 3: "complexity increment" is another strange word

choice; "increase in complexity"?

- Page 2 line 17: "is subjected to" -> "is subject to"

- Page 4 line 27: I do not understand how the the NCIt is "More than an ontology", when the rest of the sentence spells out what a good ontology should be

- Page 5 line 20: EBI's Ontology Lookup Service is not limited to OBO (flatfile) format, it supports OWL format

- Page 10 figure 1:

- `doid:Disease` should be `doid:4`

- it seems strange to use one term 'patient' from IDOMAL when there are candidates in ontologies you are already using, such as OAE

- there are existing ontologies covering gender (from multiple perspectives); the `examode.owl` file actually uses NCIt 'gender'

- Page 11 line 44: `exa:NegativeOutcome` is elsewhere referred to as `exa:NegativeResult`

- Page 12 line 1: `oae: 0001850` should not have a space

- Page 12 figure 3: HP defines 'Onset' as "The age group in which disease manifestations appear." I don't see how this can be a *subclass* of `exa:Annotation`, or a sibling to `exa:SemanticArea`.

- Page 13 line 25: HPV is referred to earlier, but this is the first time that abbreviation is spelled out.

- Page 14 figure 4: The OWL specification does not define `owl:partOf`. UBERON logically defines 'endometrium' as equivalent to "mucosa and part of some uterus", so I believe that 'endometrium' cannot be part of 'mucosa'.

- Page 15 line 4: `uberon:0001052` has label 'rectum', but ExaMode seems to be relabelling it "rectum, Not Otherwise Specified (NOS)". If this is the case, I would prefer that the authors emphasize the changes they are making. This applies to several other "NOS" terms and any other label changes.

- Page 18 table 2: There is no comparison to a "gold standard" annotation of the reports, that I can see. Figure 7 below shows that not all relevant terms are in ExaMode.

- Page 18 line 43: The fact that clinical reports were translated to English is only mentioned briefly here, but I think it requires more discussion. How effective was this translation?

- Page 19 figure 7: Not all the highlighted words are present in ExaMode -- I would like to see a discussion of this.

- Page 19 figure 8: The text of the WebVOWL graph is illegible.

- Page 20 figure 9: I am quite confused by this figure. Were there really two 'Resection' procedures? I suspect there was only one. I do not understand how 'Mild Colon Dysplasia' is the type of one outcome while being a property of the other, or how one outcome has both mild and moderate colon dysplasia.

Specific points about the examode.owl file

- I do not undertand why the various anatomical classes from UBERON and NCIt have been changed to owl:Individuals, and likewise for MONDO disease classes.

- I do not understand why MONDO:0002271 'Colon Adenocarcinoma' is an owl:Class, but MONDO:0002032 'colon carcinoma' is an owl:Individual.

- `https://hpo.jax.org/app/browse/term/HP:0003584` is not the correct identifier for ann HPO term. It should be `http://purl.obolibrary.org/obo/HP_0003584`

- The paper refers to `exa:Gender` but the OWL file uses `NCIT:C17357`.

| 2990 | This manuscript discusses the ontology engineering aspects of an ontology named ROH. It captures the entities and relations in the academic and research domains. ROH reuses concepts and relations from existing vocabularies/ontologies. Competency questions and SHACL rules are used to validate ROH. | nan | nan |

I appreciate the authors for putting in efforts to build a high quality ontology.

The following are the strengths of this submission.

1) Several terms from existing vocabularies/ontologies have been reused in ROH.

2) A good number of competency questions (CQs) have been used to validate the ontology.

3) A continuous development and integration step has been included in the ontology engineering process where the competency questions are rerun, and the ontology documentation is regenerated when there is a change to the ontology.

4) Good documentation has been provided for the ontology.

5) Permanent URLs have been used to identify the ontology resource.

6) Code is open-sourced.

7) The labels and descriptions of classes and properties are available in English and Spanish.

I have the following questions/suggestions for the authors.

1) A sample/synthetic dataset is generated and used to validate ROH. Why isn't real-world data from a university used instead? Please replace the synthetic dataset with a real-world dataset to validate the ontology.

2) The use of ontology design patterns (ODPs; http://ontologydesignpatterns.org/wiki/Main_Page) would make the ontology more modular. I would encourage the authors to explore the ODP repository and pick a few relevant ODPs to use in ROH. AgentRole ODP (http://ontologydesignpatterns.org/wiki/Submissions:AgentRole) and ActivitySpecification ODP (http://ontologydesignpatterns.org/wiki/Submissions:AgentRole) are two possibilities.

Other comments/questions.

1) As mentioned in Section 5.3, the authors consider Funding as an action. Should there be a class for an action or would a property be more suitable? To me, the latter (property) seems more appropriate, especially after looking at Fig. 7, where several classes are connected to funding through the same relation (funds). We can use roh:funds to connect these classes with the classes that roh:Funding connects to (perhaps, for example, Project?). Properties such as hasFundingID can also be added to capture the other details. Also,

on page 10, line 49 (and Eq. 1), funds to some Funding seems wrong.

2) From Section 5.3, having FundingAmount as a class is a little confusing. What are the potential instances of this class? Why can't it be a property?

3) In Table 2, is "Researcher Role" a single class? Is it meant to represent a Researcher? If so, a Research Fellowship is a type of fellowship and does not seem like a researcher or a role played by a researcher?

4) In Table 2, what is the difference between the Subject and Degree entities?

5) In Table 2, the subclasses of Internship are Predoc and PostDoc. I don't think this hierarchy is appropriate because interns are temporary whereas the other classes are full-time positions.

6) On page 10, it is mentioned that entities have categories instead of a hierarchy based on some criteria. What are these categories, and where are the criteria defined/discussed?

7) How are the rules (Eq. 1, 2, 3, ...) implemented?

8) On page 13, the range of inScheme can be either KnowledgeArea or ProjectClassification or HRClassification or FundingProgramClassification. Is it justified to use this property in four different contexts?

9) In page 17, why is hasPublicationVenue connected to a Collection? What information is captured as part of the publication venue?

10) Does Eq. (9) mean that there can only be two publication metrics for a Journal? The case when z is equal to t is not handled.

11) On page 18, hasMetric connects to Journal as well as a JournalArticle. What is the relation between a Journal and a JournalArticle? Unless one is a subclass of another, this doesn't seem right?

12) On page 20, generally, properties follow the lower camel case naming convention. For some properties, such as roh:ImpactFactorName, this is not followed.

13) Include a brief explanation of Listings 6, 7 and 8.

14) Section 6.4, why should Pellet be compiled when there are

changes to ROH? Why can't the executable of Pellet be used directly?

15) In Section 7, it was mentioned that "machine learning techniques will be applied to continuously enhance the existing contents of universities� knowledge graphs". This seems very vague. Either add a few more details or drop this line.

16) Perhaps the short Section 3 can be merged with the ontology description section.

17) Figure 1, how were the requirements gathered?

18) Table 1, how were the scenarios identified?

19) Page 9, line 44, vivo:Relationship seems to be a very general relation that can be used anywhere.

20) Please comment on the OWL 2 profile/description logic to which ROH belongs.

Typos/grammar issues

1) Page 1, Introduction, line 1, add "the" between presents and Hercules Network of Ontologies.

2) Page 1, line 42, it should be funding rather than founding.

3) Page 2, line 15, please rephrase the usage of the word "describing" here.

4) Page 2, line 20, "MA" should be "The main".

5) Page 4, line 9, the word "on" can be dropped.

6) Page 4, lines 32 and 33 should be rephrased.

7) Page 5, line 43, this line should be rephrased. "At this analysis" => "After this analysis"?

8) Page 6, line 27, it should be eg., European.

9) Page 6, line 49, "which they have participated" => with whom they have collaborated.

10) Page 6, line 50, "which are" can be removed.

11) Page 7, line 33, extend => extent

12) Page 7, line 46, at => in

13) Page 9, line 32, it should be roh:PeerReviewedArticle and not PeeReviewedArticle.

14) Page 10, where and how is the Web Annotation Data Model used? Anotation is spelt wrong in Table 4.

15) Page 11, line 7, the phrase "sparsely but just in succint remarks" should be rephrased.

16) Page 13, line 21, at => in

17) Page 20, line 13, cite => citation

18) Page 21, line 7, "Last" can be removed

19) Fig. 13, competence => competency

20) Page 26, line 27, the phrase "particularities of" can be removed.

In general, please run a grammar checker on the entire document.

| 3022 | The paper presents a a study of similarity in Wikidata and the impact that retrofitting (subsequent training of embeddings to fit with external information, in this case similarity of entity pairs in Wikidata) can have on both KG and text-based embeddings similarity. | nan | nan |

This is a very relevant topic with several interesting applications in multiple domains.

I believe the paper would be much improved by addressing the following issues:

1. Related work

There is avery relevant paper in this area that is not covered in the related work:

Lastra-D�az, Juan J., Josu Goikoetxea, Mohamed Ali Hadj Taieb, Ana Garc�a-Serrano, Mohamed Ben Aouicha, and Eneko Agirre. "A reproducible survey on word embeddings and ontology-based methods for word similarity: linear combinations outperform the state of the art." Engineering Applications of Artificial Intelligence 85 (2019): 645-665.

This is a fully reproducible paper, and producing results in the same conditions for the proposals presented in the manuscript would increase its value substantially.

Moreover, this work combines language models and KG

embeddings, but does not cover the related work that covers this overlap.

Wang, Zhen, Jianwen Zhang, Jianlin Feng, and Zheng Chen. "Knowledge graph and text jointly embedding." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1591-1601. 2014.

Xie, Ruobing, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. "Representation learning of knowledge graphs with entity descriptions." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, no. 1. 2016.

Peters, Matthew E., Mark Neumann, Robert L. Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. "Knowledge enhanced contextual word representations." arXiv preprint arXiv:1909.04164 (2019).

2. Clear definitions

The concept of retrofitting is only presented very late in the text. This is not something most readers will be familiar with and it does represent an important aspect of the work. It should be defined in the introduction to help understand the goals. A definition of KG embeddings and text embeddings is also lacking. Although these are increasingly commonplace, and soft introduction to these terms would improve the readability.

3. More focus on novel contributions

Retrofitting appears to be the more original aspect of the work. However, this is not described or analysed in depth. Results are only shown for cosine similarity of DistilRoberta embeddings.

How the pairs are built is not very well described. The word edge I believe is used to mean a pair of concepts. The selected pairs are in all likelihood quite similar (parents and siblings), and the distribution of similarity for these pairs is not studied.

"We focus our experiments on cosine similarity as a

weighting function, because we observed empirically that it consistently performs better or comparable to the other

two weighting functions." This is a pity. This is exactly what I was hoping to find in the paper. In the end, I am unsure if there is any

real advantage of using wikidata to measure conceptual similarity, or if we are simply better off just using language models.

The analysis of the quartiles is potentially quite interesting, but results are not easy to read (no table), and now the performance metric is F-measure, which is not at all clear how it is computed.

I strongly advise the authors to apply their retrofitting method in the same datasets and conditions of Lastra-Diaz et al.

4. Justifications and clarification of methodological aspects

How TopSim is computed is not clear at all. Is this the measure proposed in 10.1109/ICDE.2012.109?

Why does Composite-6 not include labels and desc?

Why is DistilRoberta used for abstract, labels, label+description and BERT-base for lexicalization?

5. The large size of wikidata is referred to multiple times, but the application was at most to a few hundred entity pairs, what are the true implications of the large size of Wikidata for similarity estimation?).

Minor: it would be better to employ the terminology defined by Lastra-Diaz et al when categorizing the different similarity metrics.

In summary, there is an interesting idea in applying retrofitting to concept similarity with KGs. However, the paper does not consider related work appropriately, which limits the value of its contributions (see Lastra-Diaz et al). It also does not afford sufficient detail in the description of the methods and choices, and could be much richer in terms of tested configuration, presented results, and discussion.

3029	The article proposes an approach to build an I4.0 benchmark dataset by using KG, called I4.0 KG, to integrate data from sensors attached to machines in a production line for the manufacture of soccer balls.	nan	nan

As a general comment, the article lacks of a concise and clear description of the key terms characteristics of the dataset. This does not help to facilitate its usage for different purposes. No information about dataset maintenance, reported usage and known shortcomings

or limitations is provided. Furthermore, the article does not demonstrate what is the advantage or the contribution of using a KG in addition to the ontological models that exist in the literature for the integration of heterogeneous data and the access to this data. I think the authors should emphasize this to make it clear to the readers. Also, the authors talk about Industry 4.0 in general but the KG they present is specific to soccer ball manufacturing. I strongly recommend the authors to dedicate a section to discuss the possibilities of adapting their approach and the proposed KG to be used in other manufacturing activities.

Here are some more specific comments:

The title of the article is not clear. I don't understand why KG appears at the end.

In the abstract:

"Our research helps the stakeholders to take timely decisions by exploiting the information embedded in the KG." - I do not see this verified in the article.

In the introduction:

Mass production was achieved in previous industrial revolutions, not in I4.0.

I4.0 is more about the use of various technologies and also the use of artificial intelligence to make better use of resources to optimize production.

The paragraph between lines 12 and 24 in page 2 is not clear. Maybe, it would be better to give the necessary definitions and then describe the differences and how they complement each other or how they are linked. Ontology and KG are not defined, please add the definitions here with the corresponding references.

The authors enumerate the contributions of the article, however I do not think they are all contributions. For example, the last one is the validation of RGOM, a model that it is not sufficiently described in the article. I do not think the validation of this model is a contribution of this article. It is more to prove that the KG built from RGOM is in fact a contribution.

In the related work:

"... Internet of Things, Internet of Services, Cyber-Physical Systems, Digital twins, ..." are not technologies. Please rephrase.

The state of the art is difficult to understand. I think the authors should work on this section and make clear the link between the existing works and what their limitations are so that it is clear how the proposed model addresses those limitations and to what extent.

In section 3:

"... acquisition and generation of the dataset." This phrase is not clear for me. Maybe, data acquisition and dataset construction ?

There is no transition between the first paragraphs of this section with subsections 3.1 ... 3.9. At the beginning two types of attributes, static and variable, are mentioned and then it goes on to describe each machine without transition. Adding a transition here would make the text more readable and aid understanding.

Regarding the random creation of the variable attribute values, the authors give a reference [17] that explains how these values are generated, however I think it would be worth to give more details about this in the article to see how accurate these values are. Also, it would be interesting to know if in the creation of these values the relationship that exists between certain properties (for example, the temperatures of the different components of a machine) is taken into account. Is it possible to represent these relationships (perhaps physicals) in the proposed KG? I think this is a key point and would allow to further enrich the data of the model.

In section 4:

"IT silos" is a very specific term, maybe give the definition or use another term like data storage.

"... usability of this data for, e.g., subsequent analyses and reasoning." - I do not understand this phrase.

Linked Open Data was never mentioned in the article before. Maybe briefly explain what it is about.

"This goal can be ..." - Which goal?

"The following describes the steps ..." -> Maybe the layers and the interaction of the different components instead of "the steps".

"... at a certain timestamps ..." -> at different timestamps.

"... unnconnected data" - What does it mean? I think it is a very general statement, it would be better to be more specific and for the authors to make it clear what they mean by "unconnected data".

"... interaction of production staff with unconnected data is very difficult." - I do not understand this phrase. It refers to access to information, interpretation, ... ?

The authors state that the RGOM model is inspired by the standards adopted by RAMI4.0. RGOM is also based on the model proposed in "Giustozzi, F.; Saunier, J.; Zanni-Merk, C. Context modeling for industry 4.0: An ontology-based proposal. Procedia Comput. Sci. 2018, 126, 675�684" and reuses other ontologies such as the Time ontology, SSN, among others. The reference [7] cites these models. The construction of the KG is not well described, it is difficult to see the link between the KG and these ontological models and how the KG is constructed from these modelsand the data. This is linked to my general comment that it is not easy to see what is the contribution of the KG with respect to using these ontological models.

I think this whole section should be rewritten and restructured to make it clear how the KG is constructed and why it is useful and necessary.

In section 5:

It is difficult to see the adaptability of RGOM through this example. The use case is very simple and does not demonstrate the usefulness of the KG. I think that the queries are too simple, maybe think about adding more complex queries that allow to see the real usefulness of the KG and the advantages it offers in terms of integration of heterogeneous data and with different temporal resolutions.

In the query of listing 3 it would be helpful to give more details about the status of the engine of a machine. I do not see the concept Status in the model, maybe it should be described how this status is obtained or what it represents (for example, if it can be seen as an

abnormal behavior). This would show that the KG offers this kind of semantic information that could be exploited by an operator and even obtain more information associated with this abnormal behavior to determine its causes.

Furthermore, maybe add third-party uses to provide evidence of the usefulness of the KG dataset. For example, a possible application (and not just say methods and tools) that can make use of KG to demonstrate its usefulness. For example, how KG could help to create suitables datasets to build machine learning models for predictive maintenance.

	Can this KG be adapted to another case that is not the manufacture of soccer balls? I think that the authors could make a discussion about this and give some hints on how to do it beyond that they do not fully validate this adaptability.		
3034	This is a joint review of Kai Eckert and Benjamin Schnabel. Benjamin Schnabel is a PhD student in the field of Digital Humanities (Jewish Studies). This revised version of the paper has been improved a lot over the previous version. All considerations have been taken into account. We therefore would recommend to accept the paper for publication.	nan	nan