Scalable Knowledge Graph Construction from Unstructured Text: A Case Study on Artisanal and Small-Scale Gold Mining

Debashis Gupta[1*], Aditi Golder[1*], Sahil Sidheekh[2], Sakib Imtiaz[3], Sarra Alaqahtani[1], Fan Yang[1], Greg Larsen[1], Miles Silman[1], Luis Fernendez[1], Robert Plemmons[1], Sriraam Natarajan[2], and V. Paul Pauca[1ξ]

[1] Wake Forest University, NC, USA
{guptd23,golda24, sarra-alqahtani, yangfan,larseng,silmanmr,fernanle,plemmons,paucavp}@wfu.edu
[2] The University of Texas at Dallas, TX,USA
{Sahil.Sidheekh,Sriraam.Natarajan}@utdallas.edu
[3] Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh
sakibimtiaz1998@gmail.com

Abstract. Constructing knowledge graphs (KGs) from unstructured text is an important yet challenging task, often requiring a great deal of manual effort and domain-specific adjustments. We propose an unsupervised, scalable method that automates both triplet extraction (subject, predicate, object) and validation using large language models (LLMs). Our framework leverages PageRank to rank external knowledge sources from the web, which are provided as contextual input to multiple LLM queries. Majority voting is then performed in the output space of these queries to ensure relationship validation. This approach offers a generalizable, domain-agnostic solution for KG construction across diverse fields. As a case study, we applied the method to Artisanal and Small-Scale Gold Mining (ASGM), constructing a knowledge graph from 1,899 triplets extracted from 9 domain-specific documents, cumulatively amounting to approximately 930 pages of unstructured text. Our framework achieves comparable performance with five baselines on a publicly available KG benchmark, and achieves over 90% accuracy on ASGM-KG, as validated by domain experts. Additionally, our framework was able to identify factual inaccuracies in popular benchmarks like Codex, highlighting the need for more reliable validation methods.

1 Introduction

Unstructured textual data represents one of the most abundant yet underutilized sources of information in the current digital era. Harnessing this vast corpus and converting it into structured, machine-interpretable formats is essential for

* → Equal contribution | ξ → Corresponding Author

advancing artificial intelligence. Knowledge graphs (KGs) have emerged as a robust framework for structuring information, facilitating applications like semantic search, recommendation systems, and automated reasoning [11,17]. Despite their potential, constructing accurate and comprehensive KGs from unstructured text remains a challenge due to issues such as scalability, reliance on domain-specific adjustments, the dependency on large annotated datasets, and the inherent ambiguity of natural language. Traditional KG construction methods, which often depend on supervised learning, are constrained by the high cost and time investment of creating annotations and their limited ability to generalize across diverse domains without substantial retraining.

To address these challenges, we propose an unsupervised, scalable framework for KG construction that leverages large language models (LLMs) for knowledge extraction and employs robust validation mechanisms. Our method eliminates the need for annotated datasets by guiding LLMs to extract subject-predicate-object (SPO) triplets. It further integrates external knowledge through a PageRank-based ranking of web sources, which provides authoritative contextual information to resolve ambiguities and improve extraction quality. To ensure reliability, we used a majority voting mechanism to validate LLM-generated outputs, reducing variance and enhancing confidence in the extracted relationships.

A second key contribution of this work is the application of our framework to construct **ASGM-KG**, a KG focused on Artisanal and Small-Scale Gold Mining (ASGM). ASGM is a critical yet underrepresented domain with profound environmental, economic, and social implications. It involves rudimentary mining practices that significantly contribute to deforestation, mercury contamination, and ecosystem degradation, particularly in tropical regions like the Amazon Basin. Collaborating with the Center for Amazonian Scientific Innovation (CINCIA) and local organizations, we curated nine domain-specific documents spanning 930 pages. Using our framework, we extracted 1,899 validated triplets to construct ASGM-KG, offering actionable insights into this complex domain.

We evaluated the framework on publicly available KG benchmarks and ASGM-KG, achieving performance comparable to or exceeding that of five baseline methods on benchmark datasets, while attaining over 90% accuracy for ASGM-KG, as validated by domain experts. Furthermore, our approach identified factual inaccuracies in existing benchmarks, such as Codex, underscoring the necessity for reliable validation mechanisms in KG construction pipelines. Thus, overall we make the following key contributions:

- An unsupervised, scalable pipeline for KG construction, leveraging LLMs and external knowledge sources to extract structured information without the need for annotated datasets.
- A novel validation mechanism employing majority voting and ensemble techniques to ensure the reliability and accuracy of the extracted knowledge.
- A case study demonstrating the application of our framework to ASGM, highlighting its environmental and social implications and providing a valuable and practical opensource resource in the form of ASGM-KG.

- Empirical validation of the framework's efficacy on both benchmark datasets and domain-specific data, showing superior or comparable performance to existing methods.

2 Background & Related Work

Early efforts in KG construction relied on rule-based methods for knowledge acquisition, involving the identification of entities, linking them to existing knowledge bases, and extracting relationships. Systems like TextRunner [30] and KnowItAll [7] utilized predefined linguistic patterns and hand-crafted rules to extract relational information from semi-structured or unstructured data at the cost of scalability and generalization.

Manual annotation of unstructured text is another approach to KG construction. Annotation tools such as WebAnno [31], INCEpTION [14], TextAnnotator [1], TeamTat [12], and SLATE [16] have been developed to facilitate this process, supporting tasks like entity recognition, relation extraction, and coreference resolution. These tools, however, remain labor intensive and time consuming for a human operator, limiting scalability with large corpora.

Recent advancements have leveraged supervised deep learning techniques to improve entity and relation extraction [23,25,29], enabling better contextual understanding and reducing noise. These models capture complex patterns in data, improving extraction accuracy compared to rule-based methods. However, they often require large amounts of annotated data for training, which is resource-intensive. Additionally, models trained on general-domain data may not perform well in specialized domains without domain-specific adaptation.

An alternative approach involves representation learning for relating textual relations to KG embeddings. Translation-based models, such as TransE [2], conceptualize the relation r in a triplet (h, r, t) as a vector translation connecting the head entity h to the tail entity t within the embedding space. Extensions like TransR [18] and TransD [13] enhance the ability to model more complex relational patterns. Attention mechanisms have also been employed to mitigate the influence of noisy data [19,27]. While these methods have advanced the field, they often remain domain-specific and may lack generalizability and robust validation without expert intervention.

Large language models (LLMs) have been recently proposed to semi-automate key tasks in knowledge graph (KG) construction, including completion, ontology refinement, and entity extraction. They can improve efficiency by generating and refining ontologies from unstructured data, enriching schemas, and extracting knowledge [22,20,4,15,8] while reducing the work of humans in the loop, particularly for validation purposes. Unlike these approaches, our method employs LLMs for extraction and validation in an unsupervised manner. Unlike expert-dependent methods, DAS leverages open-source, credible knowledge sources to ensure accuracy across diverse domains, offering a scalable approach to validating LLM-generated KGs.

You will now act as an expert in extracting **Subject Predicate Object**. Extrapolate all the available relationships from the input text named **Text**. You must follow the following instructions:

1. Every **Subject** must be a *noun*.
2. Every *pronoun* (`it`, `this`, `that`, `these`, `those`, `his`, `her`) must be strictly replaced with the *noun* that is referred to in the context by the pronoun.
3. Every **Subject** and **Object** can be a maximum of four words, and the **Predicate** strictly should be no more than three words.
4. The entities must be in order, and the order should be maintained as the text is written.
5. The order of the **Predicate** should follow the order of the text.
6. The **Predicate** should be a *verb*.
7. Try to extract the **Subject Predicate Object** in such a way that is true in your sense. Don't add any of your text in extraction.

Input: Text: {text}
Output: RDF Table. The headers of the table should be as following:
`|Subject|Predicate|Object|`

Fig. 1: Ontology Prompt used to query the LLM for extracting RDF triples.

3 Methodology

We propose a simple, scalable and unsupervised framework for constructing knowledge graphs (KGs) from unstructured text, that addresses the above discussed challenges. Our methodology comprises two main components:
1. Unsupervised Extraction of RDF Triples using LLMs: We leverage the generative capabilities of LLMs to extract RDF (Resource Description Framework) statements in the form of SPO triplets from unstructured text. By designing specific prompts and guiding the LLM with an ontology schema, we transform raw text into structured data.
2. Factual Validation via Data Assessment Semantics (DAS): We introduce an unsupervised framework called Data Assessment Semantics (DAS). This framework validates the triples by querying open-source knowledge and employing a majority voting mechanism, thereby reducing the need for domain expert intervention.

In the following subsections, we detail each component of the methodology, including algorithms and illustrative examples.

3.1 Unsupervised Extraction of RDF Triples using LLMs

Our methodology begins by extracting RDF statements in the `subject-predicate-object` format from the unstructured text corpus $\mathcal{C}$. Our framework is designed to work with any choice of an LLM, allowing for flexibility in extraction without the need for annotated training data. In this instance, we utilize `Microsoft Co-Pilot PRO` as an example, chosen for its cost-effectiveness and efficiency when tokenizing using its pre-trained weights. To guide the LLM in this task, we design a prompt based on a specified ontology schema [21], which instructs the LLM on how to identify subjects, predicates, and objects within sentences. The prompt is visualized in Figure 1. Algorithm 1 outlines the overall RDF extraction procedure. As depicted in Figure 2, the extraction workflow involves the LLM performing syntactic and semantic analysis to accurately identify

Algorithm 1 Extract RDF Triples

Require: Unstructured text corpus $\mathcal{C}$
Ensure: RDF table T with columns [Subject, Predicate, Object]

Initialize empty RDF table T
for each sentence s in $\mathcal{C}$ **do**
    $\{(s_i, p_i, o_i)\} \leftarrow$ query LLM using s as the input text in the ontology prompt
    for each triple (s_i, p_i, o_i) **do**
        if (s_i, p_i, o_i) is unique and satisfies validity constraints **then**
            Append (s_i, p_i, o_i) to T
        end if
    end for
end for
return RDF table T

subjects (nouns or noun phrases), predicates (verbs, possibly with prepositions), and objects (nouns or noun phrases). To ensure data integrity and conciseness, subject-predicate-object combinations that are contextually or logically similar to previously extracted triples are discarded, enhancing clarity and preventing duplication.

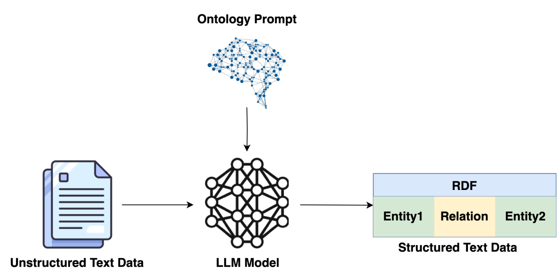

Fig. 2: Extracting RDF triples from unstructured text using an LLM.

3.2 Factual Validation via Data Assessment Semantics (DAS)

Ensuring the factual correctness of the extracted RDF triples is crucial. To achieve this without extensive manual verification, we propose an unsupervised framework called Data Assessment Semantics (DAS). The DAS framework validates the extracted RDF triples by leveraging open-source knowledge, thereby minimizing the workload on domain experts. Figure 3 illustrates the DAS framework for labeling triples as factual or non-factual and involves the following steps:

1. **Web Search and Retrieval**:For each RDF triple (s, p, o), construct a query combining the subject, predicate, and object. Submit this query to an open search engine (e.g., DuckDuckGo) and retrieve the top N relevant web pages using the search engine's API.
2. **Page Ranking and Selection**: Utilize a web page ranking tool (e.g., Open PageRank) to compute a relevance score for each retrieved page. Select the

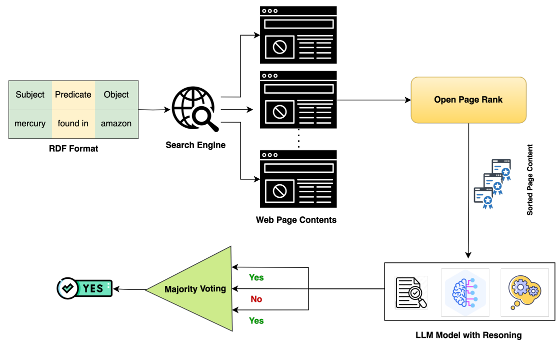

Fig. 3: Data Assessment Semantics (DAS) framework for unsupervised factual validation of RDF triples.

top K pages with scores above a specified threshold τ for further analysis. This approach ensures browser-agnostic ranking of web sources, mitigating the influence of varying internal ranking algorithms across browser, providing consistent relevant information retrieval and enhancing resource credibility.

3. **Content Summarization and Inference**:For each of the top K pages, employ the LLM to summarize the content and infer whether the RDF triple is supported by the information present. The LLM assesses the logical coherence and contextual relevance of the triple with respect to the page content.
4. **Majority Voting for Validation**:Based on the inferences from the K pages, apply a majority voting mechanism to determine the validity of the triple. If the majority indicate that the triple is valid, label it as factual; otherwise, label it as non-factual.

Algorithm 2 outlines the process of evaluating the validity of a single RDF triple using the DAS framework. In our implementation, we set the threshold τ to 7 on a relevance score scale ranging from 0 to 10, ensuring that only highly relevant pages are considered. This helps filter out less authoritative sources, improving the reliability of the validation process. This unsupervised validation process enhances the reliability of the extracted knowledge by cross-verifying each triple with multiple authoritative sources. Further, applying the DAS framework significantly reduces the number of RDF triples requiring manual verification, thus minimizing the workload on domain experts.

4 Experimental Results

To empirically demonstrate the utility of our proposed framework, we first evaluate the performance of our framework for factual validation of RDF triples against established methods using the publicly available benchmark dataset CoDEx-S [24]. We then present results from constructing a real-world knowledge graph using our framework in Section 5.

Validation of DAS Against Benchmark: Accurate classification of extracted RDF triples as factual or non-factual is critical in knowledge graph construction,

Algorithm 2 Evaluate the Validity of an RDF Triple Using DAS

Require: RDF triple (s, p, o); threshold τ; parameters N, K
Ensure: Validation result: `Valid` or `Invalid`; reasoning summary
Construct search query q by combining s, p, and o
Retrieve top N web pages $\{W_1, W_2, \ldots, W_N\}$ using search engine API
Compute relevance scores $\{r_1, r_2, \ldots, r_N\}$ for each W_i using PageRank
Select top K pages $\{W'_1, W'_2, \ldots, W'_K\}$ where $r_i \geq \tau$
Initialize validation votes $V \leftarrow \{\}$
for each page W'_i **do**
    validity$_i$ $\leftarrow$ Query LLM to to Infer if triple (s, p, o) is supported by W'_i
    Append validity$_i$ to V
end for
validation result $\leftarrow$ perform majority vote on V
return validation result

especially given the limited time that domain experts can allocate to this task. The DAS component automates this validation process. To assess its effectiveness, we compared its performance against four prevalent methods for triple classification - RESCAL, TransE, ComplEx, ConvE, and TuckER, on the CoDEx-S dataset [24], which contains 1,838 positive (factually correct) and 1,838 hard negative (factually incorrect) RDF triples. Table 1 presents the accuracy and F1 scores of DAS alongside benchmark methods. We observed that the DAS frame-

Table 1: Comparison of DAS with benchmark methods on CoDEx-S dataset.

Method	**Accuracy**	**F1 Score**
RESCAL	0.843	0.852
TransE	0.829	0.837
ComplEx	0.836	0.846
ConvE	0.841	0.846
TuckER	0.840	0.846
Ours (DAS)	0.852	0.836
Ours* (DAS With GT Correction)	**0.914**	**0.908**

work achieved an accuracy of 85.2% and an F1 score of 83.6%, comparable to benchmark methods. Note that, unlike the benchmark methods, our approach is **domain-agnostic** and does not require training with annotated datasets. To understand the misclassifications, we conducted a detailed analysis. We found that DAS correctly classified 94.4% of the negative triples but only 75% of the positive triples, according to the ground truth labels in CoDEx-S. To investigate this discrepancy, we manually reviewed the misclassified triples using expert opinions and open-source knowledge.

Our manual verification revealed that many of the positive triples misclassified by DAS were actually factually incorrect and should have been labeled as negative. Specifically, of the 448 positive triples that DAS classified as negatives, we determined that 269 were indeed negative based on authoritative sources. Similarly, among the 92 negative triples that DAS classified as positive, we found that 47 were actually positive. With the corrected ground truth labels, DAS achieved an accuracy of 91.4% and an F1 score of 90.8%. Figure 4 illustrates the validation results before and after ground truth (GT) correction. These findings demonstrate that the DAS framework not only performs comparably to existing methods but also effectively helps identify inaccuracies in benchmark datasets. Additional details and supporting evidence can be found in our GitHub repository.

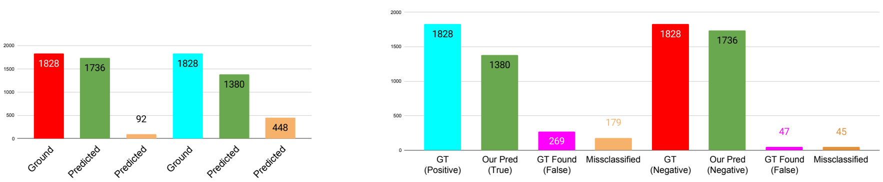

Fig. 4: Validation results of DAS on CoDEx-S dataset. Left: Results without GT correction. Right: Results after correcting GT labels, showing actual misclassifications and the number of incorrect labels identified by the framework.

5 Real-World KG: The ASGM Case Study

We applied our framework to construct a knowledge graph for Artisanal and Small-Scale Gold Mining (ASGM), demonstrating its applicability in a real-world domain. ASGM involves the extraction of gold from alluvial sediments using rudimentary methods, which often leads to significant environmental and health hazards due to mercury usage [26]. These activities contribute to deforestation, mercury contamination, and habitat destruction. While remote sensing data, such as airborne and satellite imagery, have been instrumental in analyzing large-scale environmental impacts of ASGM [3,5,6], they fail to address more nuanced questions, such as forest recovery, repeated mining at previously worked sites, and forecasting future mining activity. Furthermore, the lack of contextual information complicates efforts to inform governance and policymaking [9].

To address these challenges, we developed **ASGM-KG**, a knowledge graph specifically designed to improve understanding of ASGM's effects in tropical forests, particularly in the Amazon Basin. This initiative involved collaboration among computer scientists, ecologists, and domain experts with extensive field experience in Madre de Dios, Perú—a hotspot for both legal and illegal alluvial gold mining. Partnering with the Center for Amazonian Scientific Innovation (CINCIA) and local organizations, the team characterized critical aspects of ASGM, including deforestation, mercury contamination, and ecosystem recovery.

ASGM-KG was constructed using nine domain-specific documents curated by CINCIA experts, totaling approximately 930 pages with an average of 500 words per page. These documents are publicly accessible via our GitHub repository. Using our unsupervised RDF triple extraction method, we processed the corpus to extract structured information, yielding an initial set of 2,653 RDF triples. Figure 5 provides a visualization of a text excerpt and the corresponding RDF triples extracted. The DAS framework was then applied for factual validation,

Text Excerpt:	**Extracted RDF Triples:**
"The Amazon region is a unique environmental icon. Spanning more than a third of the South American continent, it contains the greatest share of biodiversity in the world. Further, it is home to more than 34 million human inhabitants, including some 3 million indigenous peoples. It is an invaluable source of water, food, shelter, medicines, and culture to these people of diverse origins, stretching back thousands of years. However, today the Amazon is under threat. Artisanal and small-scale gold mining is a prominent source of livelihoods and income in the region. These informal, unregulated operations make heavy use of mercury in the gold purification process, which is then released into the water and air. This mercury is carried far from the mines, irreversibly contaminating plants and animals throughout the region."	(Amazon region, is, unique environmental icon) (Amazon region, spans, more than a third of South American continent) (Amazon region, contains, greatest share of biodiversity) (Amazon region, is home to, over 34 million inhabitants) (Amazon region, includes, 3 million indigenous peoples) (Amazon region, is source of, water) (Amazon region, is source of, food) (Amazon region, is source of, shelter) (Amazon region, is source of, medicines) (Amazon region, is source of, culture) (Amazon, is under, threat) (Artisanal and small-scale gold mining, is source of, livelihoods) (Artisanal and small-scale gold mining, is source of, income) (Informal unregulated operations, use, mercury) (Mercury, is used in, gold purification process) (Mercury, is released into, water) (Mercury, is released into, air) (Mercury, is carried from, mines) (Mercury, contaminates, plants) (Mercury, contaminates, animals)

Fig. 5: Example of RDF triples extracted using our framework for ASGM. The text excerpt is on the left, and the extracted triples are on the right.

removing 754 non-factual triples and resulting in 1,899 validated triples. The final ASGM-KG comprises 1,650 unique entities and 785 unique relationships, with 43% of the entities and 29% of the relations newly discovered, as they are absent in Wikidata [28]. Manual verification by ASGM domain experts for a subset of 579 validated triples revealed a 90% matching accuracy for our framework.

To enhance accessibility and utility, ASGM-KG supports three downstream tasks: query answering, subgraph summarization, and natural language interaction. For **Query Answering**, we utilized Neo4j [10], a graph database management system, to store and query the knowledge graph. Users can explore the graph using four types of queries: retrieving all relations and objects associated with a subject node, retrieving all subjects and relations directed to an object node, finding subject-object pairs associated with a relation, and identifying relations between a specific subject-object pair. For **Subgraph Summarization**, we implemented two approaches. The first, *K-hop Summarization*, extracts subgraphs based on a specified K-hop distance, where K ranges from 1 to 5. Summaries are then generated based on the entities and relationships within each subgraph. The second approach, *Path Traversal Summarization*, identifies and summarizes paths connecting specified source and target entities, providing insights into their relationships. Lastly, we incorporated **Natural Language**

github.com/Scalable_Knowledge_Graph

Interaction using the Llama 2 70B-chat model, enabling users to query ASGM-KG using natural language. This feature eliminates the need for complex query syntax, making the system more accessible to non-technical users, and is a valuable resource for researchers, policymakers, and the public.

5.1 Ablation Study

To assess the impact of different LLMs and selection strategies on our framework's performance, we conducted an ablation study. We experimented with three LLMs: CoPilot Pro GPT-4 (Co), Llama 2 13B (L2), and Mistral 7B Instruct-v2 (M). We tested four selection strategies for our DAS framework: PageRank with Majority Voting (PR+MV), PageRank with Random Selection (PR+RS), Majority Voting without PageRank (MV), and Random Selection without PageRank (RS). We randomly selected 579 extracted triples from ASGM-KG constructed with each of the above choices and had domain experts evaluate their validity, providing a binary judgment (Yes or No) and a confidence level: `Very Confident` (90–100%), `Confident` (80–90%), `Maybe` (50–70%), or `Not Sure` (below 50%). Table 2 summarizes the validation accuracy across different models and strategies. The results indicate that Majority Voting improves validation accuracy compared to Random Selection and that PageRank is more useful for validating triplets where the expert has low confidence. These results highlight the effectiveness of incorporating Majority Voting and PageRank in our validation process. Additionally, the lower accuracy observed in less confident triples may be attributed to occasional inaccuracies in expert judgments. Notably, we observed that in cases where experts expressed low confidence, the majority of LLMs tended to agree on alternative labels. This suggests that our framework can effectively prompt experts to refine their knowledge, fostering a collaborative human-aligned AI approach.

Table 2: Validation Accuracy of Different Models and Selection Strategies Across Confidence Levels

Conf.Type	**Conf.Score**	**#Opinions**	**Co**	**L2**				**M**			
			PR+MV	**PR+MV**	**PR+RS**	**MV**	**RS**	**PR+MV**	**PR+RS**	**MV**	**RS**
Very Confident	90–100%	332	91%	94%	89%	**95%**	94%	61%	58%	62%	54%
Confident	80–90%	70	79%	87%	84%	87%	**90%**	63%	60%	56%	56%
Maybe	50–70%	81	59%	**83%**	78%	80%	80%	57%	54%	57%	57%
Not Sure	<50%	96	66%	75%	**76%**	74%	71%	46%	49%	46%	43%

6 Conclusion

In this paper, we introduced a scalable and unsupervised framework for constructing KGs from unstructured text by leveraging LLMs for RDF triple extraction and factual validation. Experimental results demonstrated the efficacy of our approach, which not only performed comparably to existing methods on the CoDEx-S benchmark dataset but also identified inaccuracies in the ground truth labels. We applied our framework to construct the ASGM Knowledge

Graph (ASGM-KG) focused on Artisanal and Small-Scale Gold Mining in tropical forests. To facilitate interaction with ASGM-KG, we implemented downstream tasks including query answering, subgraph summarization, and natural language interaction, enabling government officials, researchers, and the public to explore complex relationships within the ASGM domain, supporting informed decision-making and policy development. The high matching accuracy between DAS and domain expert evaluations in the ASGM case study further validates the effectiveness of our framework in practical applications, demonstrating the significant value of automating knowledge extraction and validation processes in domains where expert time is limited.

Acknowledgements

This research was funded in part by NASA Land-Cover and Land-Use Change Program award 80NSSC21K0309, NIH grant R01HD101246, AFOSR award FA9550-23-1-0239, ARO award W911NF2010224, and DARPA ANSR award HR001122S0039. The authors gratefully acknowledge their generous support.

References

1. Abrami, G., Mehler, A., Lücking, A., Rieb, E., Helfrich, P.: Textannotator: A flexible framework for semantic annotations. In: Proc. ACL-ISO workshop on interoperable semantic annotation. pp. 1–12 (2019)
2. Bordes, A., Weston, J., Collobert, R., Bengio, Y.: Learning structured embeddings of knowledge bases. In: Proc. of AAAI. vol. 25, pp. 301–306 (2011)
3. Caballero Espejo, J., Messinger, M., Román-Dañobeytia, F., Ascorra, C., Fernandez, L.E., Silman, M.: Deforestation and forest degradation due to gold mining in the peruvian amazon: A 34-year perspective. Remote sensing **10**(12), 1903 (2018)
4. Cohen, R., Geva, M., Berant, J., et al.: Crawling the internal knowledgebase of language models. In: Findings of the Association for Computational Linguistics: EACL 2023. pp. 1811–1824. Association for Computational Linguistics (2023)
5. Dethier, E.N., Sartain, S.L., Lutz, D.A.: Heightened levels and seasonal inversion of riverine suspended sediment in a tropical biodiversity hot spot due to artisanal gold mining. PNAS **116**(48), 23936–23941 (2019)
6. Dethier, E.N., Silman, M., Leiva, J.D., Alqahtani, S., Fernandez, L.E., Pauca, P., Çamalan, S., Tomhave, P., Magilligan, F.J., Renshaw, C.E., et al.: A global rise in alluvial mining increases sediment load in tropical rivers. Nature (2023)
7. Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Web-scale information extraction in knowitall: (preliminary results). In: Proc. of WWW. pp. 100–110 (2004)
8. Funk, M., Hosemann, S., Jung, J., et al.: Towards ontology construction with language models. In: Proceedings of the 2nd challenge on Language Models for Knowledge Base Construction. pp. 1–10. CEUR-WS.org (2023)
9. Gil, Y., Pierce, S.A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M., et al.: Intelligent systems for geosciences: an essential research agenda. Communications of the ACM **62**(1), 76–84 (2018)
10. Guia, J., Soares, V.G., Bernardino, J.: Graph databases: Neo4j analysis. In: ICEIS (1). pp. 351–356 (2017)

11. Guo, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., He, Q.: A survey on knowledge graph-based recommender systems. IEEE Trans. Knowl. Data Eng. **34**(8), 3549–3568 (2020)
12. Islamaj, R., Kwon, D., Kim, S., Lu, Z.: Teamtat: a collaborative text annotation tool. Nucleic acids research **48**(W1), W5–W11 (2020)
13. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: Proc. of IJCNLP. pp. 687–696 (2015)
14. Klie, J.C., Bugert, M., Boullosa, B., De Castilho, R.E., Gurevych, I.: The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In: Proc. of COLING. pp. 5–9 (2018)
15. Kommineni, V.K., König-Ries, B., Samuel, S.: From human experts to machines: An llm supported approach to ontology and knowledge graph construction (2024)
16. Kummerfeld, J.K.: Slate: a super-lightweight annotation tool for experts. arXiv preprint arXiv:1907.08236 (2019)
17. Liang, K., Meng, L., Liu, M., Liu, Y., Tu, W., Wang, S., Zhou, S., Liu, X., Sun, F., He, K.: A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. TPAMI (2024)
18. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proc. of AAAI. vol. 29 (2015)
19. Luo, B., Feng, Y., Wang, Z., Zhu, Z., Huang, S., Yan, R., Zhao, D.: Learning with noise: Enhance distantly supervised relation extraction with dynamic transition matrix. arXiv preprint arXiv:1705.03995 (2017)
20. Meyer, L., Stadler, C., Frey, J., et al.: Llm-assisted knowledge graph engineering: Experiments with chatgpt. arXiv preprint arXiv:2307.06917 (2023)
21. Mihindukulasooriya, N., Tiwari, S., Enguix, C.F., Lata, K.: Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text. In: ISWC. pp. 247–265. Springer (2023)
22. Pan, J., Razniewski, S., Kalo, J., et al.: Large language models and knowledge graphs: Opportunities and challenges. Transactions on Graph Data and Knowledge **1**(1), 2:1–2:38 (2023)
23. Qiao, B., Zou, Z., Huang, Y., Fang, K., Zhu, X., Chen, Y.: A joint model for entity and relation extraction based on bert. Neural Comput. Appl pp. 1–11 (2022)
24. Safavi, T., Koutra, D.: Codex: A comprehensive knowledge graph completion benchmark. arXiv preprint arXiv:2009.07810 (2020)
25. Strubell, E., Verga, P., Belanger, D., McCallum, A.: Fast and accurate entity recognition with iterated dilated convolutions. In: Proc. of EMNLP (2017)
26. Taux, K., Kraus, T., Kaifie, A.: Mercury exposure and its health effects in workers in the artisanal and small-scale gold mining (asgm) sector—a systematic review. IJERPH **19**(4), 2081 (2022)
27. Verga, P., Belanger, D., Strubell, E., Roth, B., McCallum, A.: Multilingual relation extraction using compositional universal schema. arxiv preprint. arXiv **1511** (2015)
28. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Communications of the ACM **57**(10), 78–85 (2014)
29. Wang, L., Cao, Z., De Melo, G., Liu, Z.: Relation classification via multi-level attention cnns. In: Proc. of ACL. pp. 1298–1307 (2016)
30. Yates, A., Banko, M., Broadhead, M., Cafarella, M.J., Etzioni, O., Soderland, S.: Textrunner: open information extraction on the web. In: Proc. of NAACL-HLT. pp. 25–26 (2007)
31. Yimam, S.M., Gurevych, I., De Castilho, R.E., Biemann, C.: Webanno: A flexible, web-based and visually supported system for distributed annotations. In: Proc. of ACL: System Demonstrations. pp. 1–6 (2013)