

# A KNOWLEDGE GRAPH MODELING APPROACH FOR AUGMENTING LANGUAGE MODEL-BASED CONTRACT RISK IDENTIFICATION

Chunmo Zheng<sup>1</sup>, Yinqiu Tang<sup>2</sup>, and Xing Su<sup>1</sup>

<sup>1</sup> Zhejiang University, Hangzhou, China

<sup>2</sup>PowerChina Huadong Engineering Corporation Limited, Hangzhou, China

## Abstract

Contract risk identification is essential for preventing disputes and losses in construction industry. Large language models (LLMs) have impacted various natural language processing tasks, offering a promising avenue for automating contract review without extensive data processing and feature engineering. However, LLMs still has difficulty in recalling facts while generating knowledge-grounded analysis, especially when related to complex domain knowledge. This paper introduces a Knowledge Graph (KG) modeling approach to enhance the LLM-based automated contract risk identification. A case study demonstrates that our approach exhibits enhanced performance on risk identification tasks compared to non-augmentation scenario.

## Introduction

Legal issues, such as disputes, claims, and litigation, frequently occur in construction industry, due to the growing size and complexity of construction projects. The use of natural language in contracts has been identified as the leading cause of such issues (Zait & Zarour, 2018). Specifically, semantic vagueness and ambiguities in contracts can lead to disagreements and misunderstandings between parties, resulting in conflicts or disputes (Mahfouz et al., 2018).

Natural Language Processing (NLP) techniques offers a promising tool to automate contract text processing, thus reducing human errors and increase efficiency. NLP has demonstrated its potential in various automated text processing tasks within the construction industry, including automated contract reviewing (Lee et al., 2020), automated compliance checking (Salama & El-Gohary, 2016) and similar case retrieval (Zou et al., 2017). These techniques can be broadly categorized into rule-based, machine learning-based and deep learning-based methods. However, two limitations have been frequently highlighted in the current literature. First, existing methods rely on sufficient high-quality annotated data to perform the domain-specific tasks, which is highly data-dependent and lack in scalability. Second, current approaches are facing the criticism of only predicting fragmented risk labels without an interpretable risk reasoning.

To mitigate the aforementioned problems, researchers have proposed using large language models trained on a massive amount of unstructured data to facilitate the legal and contractual reasoning tasks. For example, LLMs have begun to reshape the practice of legal services among professionals (Fernandez, 2023). Many argues that the model's ability to learn new tasks from limited data would significantly reduce the human labor in data annotating and feature engineering. However, it still faces many challenges when adapting LLMs to downstream contract risk reasoning. They may struggle with domain-specific knowledge and may not be able to provide explanations from expert perspective, along with factual errors.

Existing works have shown that retrieving structured triple-formed knowledge from KGs can significantly improve LLM's performance on knowledge-intensive reasoning. KG is a data structure designed to accumulate and represent real-world knowledge, where nodes represent entities of interest and whose edges represent various relations between these entities (Hogan et al., 2021). The explicit knowledge representation of KGs can be provided as external knowledge to guide LLMs towards more robust and interpretable knowledge reasoning process (Yang et al., 2024).

Motivated by the identified challenges and existing KG-augmented LLMs solutions, this paper proposes a nested knowledge graph (KG)-enhanced LLM method for automated contract risk identification. A comprehensive framework of developing the ontological layer and instance layer of contract KG is introduced. The nested KG can capture the intricate and interconnected relation within contract clauses, which provide structured and accurate external knowledge for LLMs. Our approach shows improved interpretability of the generated content, and also sheds a light on the integration of LLMs and KGs in automated contract review systems and other knowledge-intensive domain reasoning tasks.

## Literature Review

### NLP-assisted contract management and the role of prior knowledge

NLP techniques have been widely used in addressing legal issues in construction projects (Hassan et al., 2021).

In contract review, many researchers perform the detection of requirement clauses or poisonous clauses. Lee et al. (2019, 2020) use NLP and semantic syntactic rules in both risk-prone clauses extraction and contractor-friendly clauses detection, achieving the F-score of 81.8% and 80%. Machine learning-based NLP is employed to detect risk-prone paragraphs in contract, achieving an accuracy of 94% (Chakrabarti et al., 2018). Additionally, many studies have also used machine learning algorithms to classify contract clauses into different categories to facilitate contract review process. The categories include requirement or non-requirement clauses, different categories of construction projects or different categories of risks (Candaş & Tokdemir, 2022; Hassan & Le, 2021).

Turning to automated regulation text processing, automated compliance checking aims at detecting the violation with construction laws or regulations. Semantic text classification in unstructured provisions is the first step to perform the automated compliance checking task (Beach et al., 2020). Salama et al. (2016) used a hybrid approach combining semantic analysis and machine learning to classify clauses, achieving a recall of 100%. Zhou and El-Gohary (2016) introduced a rule-based NLP method to classify environmental requirements, alongside a machine learning-based approach for regulatory codes classification.

These studies reveal that NLP techniques have achieved a promising performance in the contract document analysis. However, owing to the unique and complex nature of contract terminologies and semantic patterns, it is difficult for generic NLP models to produce equally reliable performance in construction contract domain without a proper adaptation (Hassan et al., 2021). A few recent studies have attempted to settle these issues by integrating domain-knowledge into the NLP models with the help of semantic representations such as taxonomy, ontology and KG (Lee et al., 2019; Xu & Cai, 2019; J. Zhang & El-Gohary, 2016). These studies have shown an impressive performance, highlighting the importance of integrating domain-knowledge in the future NLP-assisted contract analysis.

### Knowledge representation of domain knowledge and the issue of complex knowledge modeling

Semantic web technologies enable the representation of machine-readable data on the web (Kumar, 2019). It is widely adopted in the Knowledge Representation (KR) of domain knowledge by defining domain concepts, enhancing domain information integration and performing logical reasoning (Li et al., 2022). Ontology is one of the fundamental concepts in semantic web technology. It provides a conceptual model to describe a set of concepts in a domain (S. Zhang et al., 2015). Many researchers have developed construction domain ontology for efficient knowledge management, compliance checking, risks or conflicts detection (Zhong et al., 2019; Z. Zhou et al., 2016).

Ontology can function as formal representation to create knowledge graphs (KGs). Recent years, KG has been

appeared as a major trend in KR technique to serve many industrial applications (Fang et al., 2020). It presents knowledge in the form of labeled directed graph. Each entity is considered as a node and they are linked via edges which represent relations between entities. The basic unit of KG can model binary semantics with RDF triple (Duan et al., 2017).

In real world knowledge of many domains, especially contract domain, there are always additional information conveying conditional, temporal or provenance information that are beyond the modeling ability of a triple. Modeling and extracting such metadata are beneficial for efficient domain knowledge management. Semantic legal metadata extraction is also crucial for interpreting legal provisions (Sleimi et al., 2019, 2021).

Data modeling solutions for RDF metadata include standard reification, singleton property and RDF-star (Hartig & Thompson, 2021). They are able to represent additional contextual information attached to individual triples (Orlandi et al., 2021). The metadata representation in the graph structure have also been explored. Temporal KG and event KG associate triples with time or site hyper-edges, where an event triple is considered as a semantic unit (Lv et al., 2022; Park et al., 2022).

### Contract KG modeling approach

Motivated by the challenges of representing complex knowledge in contract, we introduce the nested KG framework. This framework defines a triple in knowledge graphs as head and tail nodes connected by a relationship edge, while allows nodes to appear as either a single node or a fact node. A fact node is a nested entity which contains a triple inside. This allows a triple to be treated as a node, so that further constraints or statement can be link to the triple. The symbolic definition of this framework is given in Table 1.  $h, r, t$  denotes the head entity, relation and tail entity respectively.  $E, F, R$  refers to the collection of single entity, fact (nested entity) and relation. Single entities and facts can be interlinked with each other to form nested triples, namely entity-to-entity, entity-to-fact and fact-to-fact triple.

Table 1: Definition of nested triples

Triple	Definition
$T_{E2E}$	$T_{E2E} = \{(h, r, t)   h, t \in E, r \in R\}$
$T_{E2F}$ $/T_{F2E}$	$T_{E2F} = \{(h, r, t)   h \oplus t \in \{E, F\}, r \in R\}$
$T_{F2F}$	$T_{F2F} = \{(h, r, t)   h, t \in F, r \in R\}$

By expanding the concept of entities and triples in knowledge graphs to multi-layer nested structure, the knowledge representation framework can capture deeper semantics and interconnected facts. When dealing with the complex relations in contract context such as conditional, temporal and compound relations, our framework is able to incorporate not only the scattered concepts, but relations among the concepts.

Principally, the nested structure can be infinitely nested, in order to represent more complex semantics. However, it is important to note that there exists the trade-off between expressivity and scalability of knowledge graph, where the more complex nested layer will limit the graph scalability.

### Ontology layer development

In this section, a contract ontology based on the nested knowledge graph framework is developed to represent contractual knowledge. As shown in Figure 1, the ontology provides a hierarchical conceptual schema for modeling contract semantics.

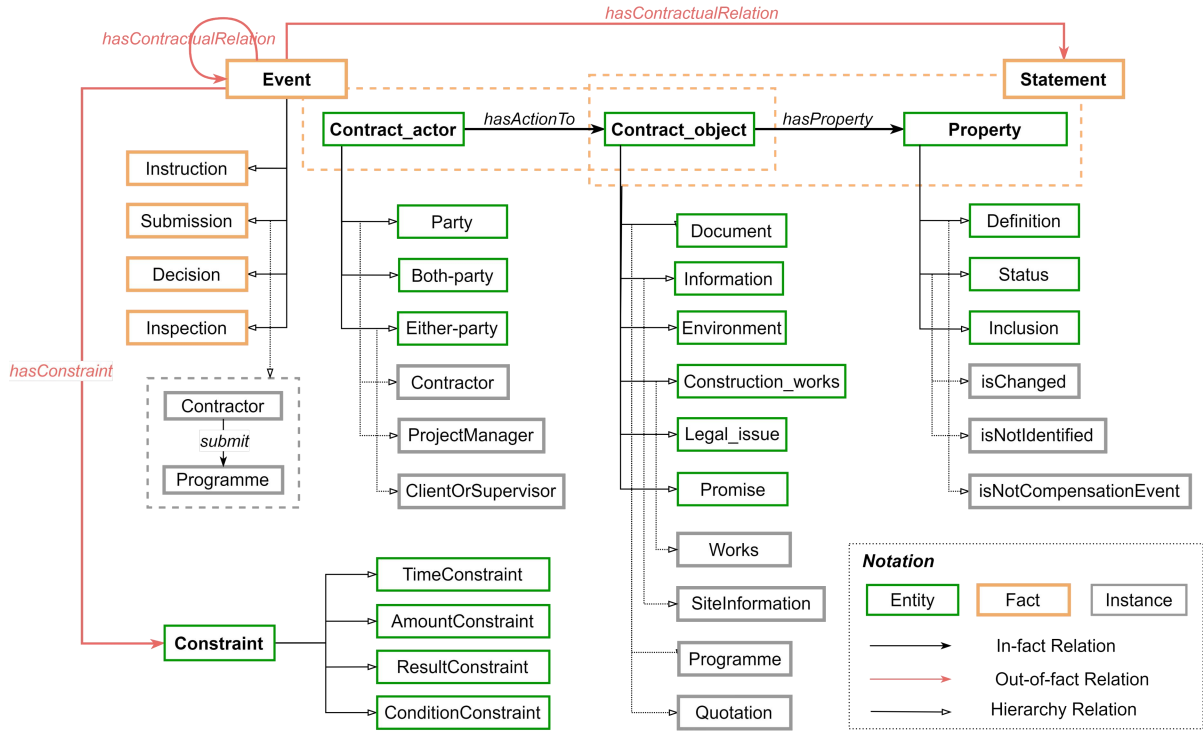


Figure 1: Ontological layer of nested contract KG

The core entities in the ontological layer of nested contract knowledge graph contains “Contract\_actor”, “Contract\_object”, “Property” and “Constraint”. “Contract\_actor” and “Contract\_object” class can be joined by the “hasActionTo” relation to form the basic “Event” class fact: <Contract\_actor, hasActionTo, Contract\_object>. For example, <Contractor, submit, Programme> is an instance of “Event” class, and more specifically, it is under the “Submission” subclass. Similarly, the description of the property of certain product is called a “Statement”. The fact <Programme, hasProperty, isNotIdentified> is an instance of “Statement” class.

In the contract clauses, a “Event” always has many constraints. The “Constraint” class can be attached to the “Event” class as <Event, hasConstraint, Constraint>. In real instances, the relation and tail entity should be replaced by specific constraint types such as <Event, hasTimeConstraint, beforeCompletionDate>. Finally, we have the “hasContractualRelation” edge to connect a Event and a Statement, or connect two Events. The

contractual relations can either be conditional relations or temporal relations in the contract. Based on this, we can extract the contract rules as, for example, <Statement, if-then, Event> or <Event1, withinPeriodOf, Event2>.

### Instance layer extraction

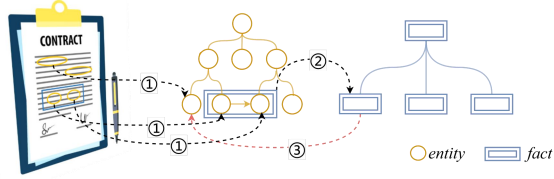
For representing the contract knowledge graph instances, RDF-star language is chosen for its efficiency in representing nested data structures efficiently and directly, without introducing redundant nodes.

Figure 2 illustrates the procedure of constructing nested knowledge graph from sample construction contract

clause. First, entity extraction is performed by linking the identified entities to the predefined ontology classes. For example, “Contractor” and “Project Manager” belongs to Party class, while “Proposal” belongs to Document class. Accordingly, two Fact is recognized as (Contractor, make, Proposal) and (Project Manager, accept, Proposal), denoting two events in the contract. Finally, temporal relation “within two weeks of” is used to link the two facts and form the knowledge representation. The final RDF-star representation is shown in the rectangular box.

Following the ontology and construction workflow defined above, we can also develop automated approach to facilitate the knowledge graph construction. Given the in-context learning ability of LLMs, recent studies proposed using ontology prompting to generate structured knowledge representation. Motivated by such efforts, our instance development method adopts a semi-automated construction following knowledge

representation from unstructured text<sup>1</sup>. This can reduce the intensive human efforts from constructing a knowledge graph to an acceptable level of manual intervention.



Contract clause: **Within four weeks of the Contractor making the proposal the Project Manager accepts the Contractor's proposal.**

① **entity identification**  
Entity1:Contractor (Party),  
Entity2:Project Manager (Party);  
Entity3:Proposal (Document)

② **fact identification**  
Fact1: (Contractor, make, Proposal);  
Fact2:(Project Manager, accept, Proposal)

③ **relation linking**  
→ (fact1, within2weeksOf, fact2)

```
@prefix ckg:<http://ContractKG/ckg#> .
<<ckg:ProjectManager ckg:accept ckg:Proposal>>
ckg:within2weeksOf <<ckg:Contractor ckg:make ckg:Proposal>> .
```

Figure 2: Schematic diagram of constructing nested KG from contract

### Nested Contract KG query

The nested contract KG query allows the entity-level knowledge retrieval from the constructed domain knowledge database. This section will briefly introduce the construction of domain knowledge base and the contract knowledge query. The ontological layer of contract KG was created in Protégé, and the instance layer was stored in RDF-star language as a turtle file. The classes and instances were then imported into graph database such as GraphDB or Neo4j. Therefore, the query can be performed through SPARQL or Cypher language.

In practice, standard form of construction contracts (SFCCs) serves as a standard for regulating the responsibility and roles for stakeholders. They are the important knowledge for understanding and reviewing the contracts. We present sample queries of the SFCC KG as in Figure 3. In the first query in (a), contractor may want to know all of the time constraints for submitting documents. Through graph database, the documents and corresponding time period for submitting will be returned both in RDF and graph view. We can also explore interlinked concepts in the graph database. As shown in (b), we can perform a more complex query related to conditional relation of two “Event” class. A triple will be retrieved to indicate what a party should do in this case.

<sup>1</sup> <https://bratnic-tomaz.medium.com/constructing-knowledge-graphs-from-text-using-openai-functions-096a6d010c17>

```
Q (TimeConstraint): What are the time constraints for documents that Contractor is responsible for submitting?
PREFIX ckg:<http://ContractKR/ckg#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>

SELECT ?o ?b WHERE {
BIND (<<ckg:Contractor ckg:submit ?o>> AS ?a)
?a ckg:hasTimeConstraint ?b .
}
```

(a)

```
Q (Behavior): What should parties do if quotation for an acceleration is accepted?
PREFIX ckg:<http://ContractKR/ckg#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>

SELECT ?c WHERE {
BIND (<<ckg:QuotationForAnAcceleration ckg:hasProperty ckg:isAccepted>> AS ?a)
?a ckg:if-then ?c .
}
```

(b)

Figure 3: Demonstration of query in contract KG

### Augmenting LLMs with contract KG on automated contract review

In this section, we propose the overall framework to integrate nested knowledge graph in automated contract review using natural language prompting.

Figure 4 illustrate our contract KG-enhanced LLM contract review framework. The aim is to retrieve external knowledge in the nested contract KG for augmenting contract analysis task. Given a contract clause  $c$ , a set of natural language query  $Q = \{q_1, q_2, \dots, q_n\}$  will be generated, and subsequently transformed into SPARQL query  $S = \{s_1, s_2, \dots, s_n\}$ . External explicit knowledge  $K = \{k_1, k_2, \dots, k_m\}$  will be retrieved in RDF-star format and feed back to LLMs. The final prompt  $P = \{r|c, k\}$  is designed to generate corresponding risk analysis result with the initial contract clause input and the retrieved knowledge.

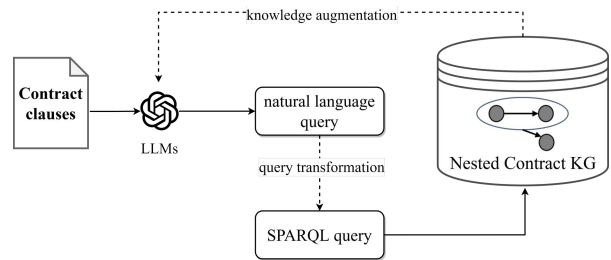


Figure 4: Contract KG-augmented LLM contract review pipeline

Specifically, in the first query generation stage, the language model is prompted to analyze the review task and output relevant queries based on the given contract clauses. The prompt is designed for asking the questions such as “What terms in this clause need further clarification?” and “What knowledge about the clause is required to determine the risk?”. The output should be querying property of certain terms, constraints of contract

events as external knowledge of the contract context. In this stage, we aim to stimulate the in-context learning ability of LLMs for identifying what external knowledge is needed in the context, to bridge the knowledge gap with experts. In the query transformation stage, the natural language to SPARQL query translation requires LLMs to map the concept in natural language query to classes in the defined knowledge graph. Multi-hop query can also be included in order to respond to the natural language query. After the knowledge retrieval from nested contract KG, the triple-formed knowledge and the initial contract clause will be input for prompting LLMs to generate the final result. This whole process performs in a retrieval-augmented fashion, providing external knowledge for the automated contract risk identification.

## Case study

In this section, we validated our approach on contract clauses collected from international construction projects. The contract was thoroughly reviewed by contract experts, and the sample clauses-review pair was given as the gold standard, as shown in Table 2. The sample clause was input for risk identification and analysis in the standard LLM review (baseline) and KG-enhanced LLM contract review pipeline (our method), respectively. In the baseline setting, only contract clause was input for LLMs to identify the potential risk, while our method also provided the final output according to the process described in Figure 5. The evaluation process was manually conducted by domain experts, measuring both the risk label and the accuracy of the risk analysis.

The baseline output and the process of our approach is shown in Table 3. All the content was generated using gpt-3.5-turbo, which is a natural language understanding and generating model provided by OpenAI<sup>2</sup>.

Table 2: Contract clause review case overview

Project clause& review demonstration
Contract clause: (a) Contract clause: The amount of the Advance Payment shall be twenty five (25%) of the Supply Price and shall be amortised proportionately against amounts recovered through interim applications for payment throughout the progress of the Supply, starting with the first Interim Payment after the Order Date. The currency shall be the same as the Supply Price.
(b) The Employer shall pay the Advance Payment after: (i) receiving the Performance Bond in accordance with Clause 4.3(b), and (ii) receiving the Advance Payment Guarantee in accordance with Clause 4.3(b), provided, however, the Advance Payment shall be made within fifteen (15) days of the later of (i) and (ii).
(c) The Supplier shall ensure that the Advance Payment Guarantee is valid and enforceable until the entire Advance Payment has been adjusted by the Supplier, but its amount shall be progressively reduced by the amount adjusted through proportional deductions from Interim Payment amounts paid for the Equipment under Clause 11.5 (Schedule of Payments) ...

Risk review: No risk. The amount of Advance Payment is quite generous to the Contractor, which is much higher than the common practice of 10%. Requiring Performance Bond or Advance Payment Guarantee prior to paying the Advance Payment is reasonable practice to protect the interest of the Employer. The requirements for proportional deductions from Interim Payment and validity period are also common practice, which reminds the Contractor of updating the guarantee if necessary. It would be better if the contract could stipulate the Contractor's right when the release of the Advance Payment is delayed by the Employer.

Table 3: Knowledge-enhanced LLM contract review

Automated risk analysis demonstration
<b>Baseline output</b>
Risk. The clause introduces financial risk for the Contractor because the Advance Payment is tied to the provision of a Performance Bond and an Advance Payment Guarantee, which can be costly and administratively burdensome. Additionally, the Advance Payment must be repaid through deductions from interim payments, which could affect the Contractor's cash flow. The obligation to maintain the Advance Payment Guarantee until the Advance Payment is fully adjusted adds further financial strain and administrative responsibility on the Contractor.
<b>Natural language query</b>
Explain Advance Payment Guarantee. [Q1]
What is the condition of making advance payment? [Q2]
<b>SPARQL query</b>
Q1:
SELECT ?p ?actor1
WHERE {
ckg: advancePaymentGuarantee ckg:hasProperty ?p .
?actor1 ckg:submit ckg: advancePaymentGuarantee }
Q2:
SELECT ?actor2 ?c1 ?c2 WHERE {
BIND (<<?actor2 ckg:make ckg:advancePayment >> AS ?e)
?e ckg:hasTemporalConstraint ?c1 .
?e ckg:hasConditionalConstraint ?c2 }
<b>Retrieved knowledge from nested KG</b>
A1:
ckg: Employ ckg:submit ckg: advancePaymentGuarantee .
ckg: advancePaymentGuarantee ckg:hasProperty ckg: amountequalToAdvancePayment .
ckg: advancePaymentGuarantee ckg:hasProperty ckg: currenctyequalToAdvancePayment .
A2:
<<ckg: Employer ckg:receive ckg:PerformanceBond_and_advancePaymentGuarantee >> ckg:within14days << ckg: Employer ckg:make ckg:advancePayment>>
<b>KG-augmented output</b>
No risk. The clause matches the standard provision regarding the advance payment being conditional on receiving the Performance Bond and Advance Payment Guarantee. The payment timeline of 15 days, as opposed to 14 in the standard form, is slightly longer but does not significantly alter the contractor's obligations or risks. The requirement for the

<sup>2</sup> <https://platform.openai.com/docs/models/overview>

---

Advance Payment Guarantee to be valid until the advance is fully adjusted is also standard, ensuring the employer's risk is mitigated without imposing unbalanced obligations on the contractor.

---

As shown in Table 3, the baseline output using standard prompting stated that the contractual risk lies in the precondition of providing performance bond and advance payment guarantee before advance payment, posing an unbalanced obligation to the contractor. However, it is an unfounded claim as employer make payment based on receiving advance payment guarantee. Such hallucination arises due to the absence of knowledge related to relevant contract terms and provisions. In the retrieval process, the regulation of advance payment guarantee and constraints of making advance payment was retrieved as the standard provision. After knowledge-augmentation, the model answers “No risk”, with an explanation that objectively analyze the contract provision in comparison to the standard provision. The answer aligns with the expert review and showcases a more coherent and interpretable reasoning in the perspective of a contract expert.

## Discussion

In this paper, we introduce the automated contract review method enhanced with contract knowledge graph. This method suggests integrating LLMs and KG to achieve a more interpretable and trustworthy automated contract review process. Although it is a feasible solution, there are some unsolved challenges and future promising directions that we would like to discuss here.

Firstly, there are also other approaches to augment LLMs, such as retrieving knowledge from websites and tools. They can enhance LLMs with real-time information or mathematics ability, which can be combined with domain knowledge graph for comprehensive knowledge augmentation.

Secondly, the current process is conducted semi-automatically, while further study is required for higher level and more accurate automation of ontology construction and query mapping. This might involve ontology learning and searching for the relevant reasoning path in KGs. It is noted that although achievements have been made in ontology extraction, current methods still struggle with complex multi-layer extraction, making human intervention for nested knowledge graph construction inevitable at the moment. The trade-off between expressivity and scalability should also be considered. More accurate and comprehensive knowledge representation comes with higher human intervention.

Thirdly, the current evaluation metrics does not consistently align with expert evaluations concerning the accuracy and coherence of risk reasoning output. This is because the existing benchmarks focus on the tasks where model learn by fine-tuning on task-specific data. Under the LLM era, constructing a benchmarks for construction contract risk analysis will be of great importance.

## Conclusion

This paper presents a novel framework for augmenting large language models with structured domain knowledge to improve automated contract risk identification. A knowledge graph modeling approach is developed to capture the complex and interconnected relationships within contracts, along with the contract ontology and RDF-star representation. A framework that integrates knowledge from knowledge graph into the LLM reasoning pipeline using natural language prompting is also proposed. The case study demonstrates improved performance with our knowledge augmentation method on automated contract risk identification task.

## Acknowledgement:

This research was funded by the National Natural Science Foundation of China (grant no. 71971196) and China Postdoctoral Science Foundation (No. 2023M743307).

## References

- Beach, T. H., Hippolyte, J.-L., & Rezgui, Y. (2020). Towards the adoption of automated regulatory compliance checking in the built environment. *Automation in Construction, 118*, 103285. <https://doi.org/10.1016/j.autcon.2020.103285>
- Candaş, A. B., & Tokdemir, O. B. (2022). Automating Coordination Efforts for Reviewing Construction Contracts with Multilabel Text Classification. *Journal of Construction Engineering and Management, 148*(6), 04022027. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002275](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002275)
- Chakrabarti, D., Patodia, N., Bhattacharya, U., Mitra, I., Roy, S., Mandi, J., Roy, N., & Nandy, P. (2018). Use of Artificial Intelligence to Analyse Risk in Legal Documents for a Better Decision Support. *TENCON 2018 - 2018 IEEE Region 10 Conference*, 0683–0688. <https://doi.org/10.1109/TENCON.2018.8650382>
- Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., & Lin, Z. (2017). Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 327–332. <https://doi.org/10.1109/SERA.2017.7965747>
- Fang, W., Ma, L., Love, P. E. D., Luo, H., Ding, L., & Zhou, A. (2020). Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Automation in Construction, 119*, 103310. <https://doi.org/10.1016/j.autcon.2020.103310>
- Fernandez, J. V. (2023). Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State. *Indiana Journal of Global Legal Studies, 30*, 65.
- Hartig, O., & Thompson, B. (2021). *Foundations of an Alternative Approach to Reification in RDF*. <http://arxiv.org/abs/1406.3399>
- Hassan, F. ul, & Le, T. (2021). Computer-assisted separation of design-build contract requirements to support subcontract drafting. *Automation in Construction, 122*, 103479. <https://doi.org/10.1016/j.autcon.2020.103479>
- Hassan, F. ul, Le, T., & Lv, X. (2021). Addressing Legal and Contractual Matters in Construction Using Natural Language Processing: A Critical Review. *Journal of Construction Engineering and Management, 147*(9), 03121004. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002122](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002122)
- Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Gayo, J. E. L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngomo, A.-C. N., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge Graphs. *ACM Computing Surveys, 54*(4), 1–37. <https://doi.org/10.1145/3447772>
- Kumar, V. K. (2019). Knowledge Representation Technologies Using Semantic Web. In *Web Services: Concepts, Methodologies, Tools, and Applications* (pp. 1068–1076). IGI Global.
- Lee, J., Ham, Y., Yi, J.-S., & Son, J. (2020). Effective Risk Positioning through Automated Identification of Missing Contract Conditions from the Contractor's Perspective Based on FIDIC Contract Cases. *Journal of Management in Engineering, 36*(3), 05020003. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000757](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000757)
- Lee, J., Yi, J.-S., & Son, J. (2019). Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *Journal of Computing in Civil Engineering, 33*(3), 04019003. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000807](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000807)
- Li, X., Yang, D., Yuan, J., Donkers, A., & Liu, X. (2022). BIM-enabled semantic web for automated safety checks in subway construction. *Automation in Construction, 141*, 104454. <https://doi.org/10.1016/j.autcon.2022.104454>
- Lv, X., Shi, J., Cao, S., Hou, L., & Li, J. (2022). Triple-as-Node Knowledge Graph and Its Embeddings. In A. Bhattacharya, J. Lee Mong Li, D. Agrawal, P. K. Reddy, M. Mohania, A. Mondal, V. Goyal, & R. Uday Kiran (Eds.), *Database Systems for Advanced Applications* (Vol. 13245, pp. 107–121). Springer International Publishing. [https://doi.org/10.1007/978-3-031-00123-9\\_8](https://doi.org/10.1007/978-3-031-00123-9_8)
- Mahfouz, T., Kandil, A., & Davlyatov, S. (2018). Identification of latent legal knowledge in differing site condition (DSC) litigations. *Automation in Construction, 94*, 104–111.
- Orlandi, F., Graux, D., & O'Sullivan, D. (2021). Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF. *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, 233–240. <https://doi.org/10.1109/ICSC50631.2021.00049>
- Park, N., Liu, F., Mehta, P., Cristoforo, D., Faloutsos, C., & Dong, Y. (2022). EvoKG: Jointly Modeling Event Time and Network Structure for Reasoning over Temporal Knowledge Graphs. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 794–803. <https://doi.org/10.1145/3488560.3498451>
- Salama, D. M., & El-Gohary, N. M. (2016). Semantic Text Classification for Supporting Automated Compliance Checking in Construction. *Journal of Computing in Civil Engineering, 30*(1), 04014106. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000301](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000301)

- Sleimi, A., Ceci, M., Sannier, N., Sabetzadeh, M., Briand, L., & Dann, J. (2019). A Query System for Extracting Requirements-Related Information from Legal Texts. *2019 IEEE 27th International Requirements Engineering Conference (RE)*, 319–329. <https://doi.org/10.1109/RE.2019.00041>
- Sleimi, A., Sannier, N., Sabetzadeh, M., Briand, L., Ceci, M., & Dann, J. (2021). An automated framework for the extraction of semantic legal metadata from legal texts. *Empirical Software Engineering*, 26(3), 43. <https://doi.org/10.1007/s10664-020-09933-5>
- Xu, X., & Cai, H. (2019). Semantic Frame-Based Information Extraction from Utility Regulatory Documents to Support Compliance Checking. In I. Mutis & T. Hartmann (Eds.), *Advances in Informatics and Computing in Civil and Construction Engineering* (pp. 223–230). Springer International Publishing. [https://doi.org/10.1007/978-3-030-00220-6\\_27](https://doi.org/10.1007/978-3-030-00220-6_27)
- Yang, L., Chen, H., Li, Z., Ding, X., & Wu, X. (2024). Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling. *IEEE Transactions on Knowledge and Data Engineering*, 1–20. <https://doi.org/10.1109/TKDE.2024.3360454>
- Zait, F., & Zarour, N. (2018). Addressing Lexical and Semantic Ambiguity in Natural Language Requirements. *2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT)*, 1–7. <https://doi.org/10.1109/ISIICT.2018.8613726>
- Zhang, J., & El-Gohary, N. M. (2016). Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *Journal of Computing in Civil Engineering*, 30(2), 04015014. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000346](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346)
- Zhang, S., Boukamp, F., & Teizer, J. (2015). Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction*, 52, 29–41.
- Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, 101, 17–31. <https://doi.org/10.1016/j.autcon.2018.12.013>
- Zhou, P., & El-Gohary, N. (2016). *Automated Extraction of Environmental Requirements from Contract Specifications*.
- Zhou, Z., Goh, Y. M., & Shen, L. (2016). Overview and Analysis of Ontology Studies Supporting Development of the Construction Industry. *Journal of Computing in Civil Engineering*, 30(6), 04016026. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000594](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000594)
- Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Automation in Construction*, 80, 66–76. <https://doi.org/10.1016/j.autcon.2017.04.003>