Constraint-Aware Ontology Engineering for Information Extraction from Financial Contracts

Ryoma Kondo[1,2,*,†], Shamik Kundu[3,†], Muneaki Imai[3,†], Kyosuke Tomoda[3,†], Yuki Takazawa[1] and Ryohei Hisano[1,2]

[1] *Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[2] *The Canon Institute for Global Studies, ShinMarunouchi Building 5-1 Marunouchi 1-chome, Chiyoda-ku, Tokyo 100-6511 Japan*

[3] *Trust Co.,Ltd., THE SHORE Nihonbashi Kayabacho 5F, 1-2 Nihonbashi Hakozakicho, Chuo-ku, Tokyo 103-0015, Japan*

Abstract

The automation of business processes is advancing rapidly, particularly in domains governed by complex rule-based documentation such as financial contracts. These documents are highly structured and semantically rich, which makes them well suited for modeling with formal ontologies. Although large language models offer promising capabilities for information extraction, their effectiveness is limited by the challenge of designing consistent prompts across contract types due to a lack of standardized semantic definitions. In this paper, we explore how embedding ontological structures and constraint specifications into prompts can improve the accuracy and reusability of information extraction systems, using confirmation notices for investment transactions as a case study. Our findings show that incorporating semantic constraints into prompts improves performance in language model-based information extraction, highlighting the potential of combining Semantic Web technologies with language models to support accurate and maintainable information extraction from financial contracts.

Keywords

Ontology-based Information Extraction, Semantic Constraints, Ontology Engineering, Large Language Models,

1. Introduction

The demand for business process automation is rapidly increasing, particularly in the domain of legal contracts [1, 2, 3, 4]. These contracts are typically highly structured and governed by formal legal and domain-specific rules, which makes them well suited for ontological modeling. As a result, the Semantic Web community has long viewed legal contracts as a promising area for applying ontologies to formally define the structure and semantics of relevant information [5, 6, 1, 7, 8]. Such modeling aids the development of highly reusable and interoperable information extraction specifications.

Recently, the advent of large language models (LLMs) has renewed interest in ontology-based approaches [9], given their potential to support structured information extraction from contracts. For example, some recent work applies an ontology-guided prompting framework to extract key fields from financial documents using LLMs [10, 11]. To guide such efforts, it is important to empirically assess how ontologies and constraints influence extraction accuracy.

Financial contracts, such as investment agreements, derivatives, and loans, are a compelling domain for Semantic Web technologies [1, 12]. Though structurally and terminologically diverse, they often contain well-defined information (e.g., amounts, dates, parties) governed not only by intra-property constraints [13, 14], but also by inter-property constraints [15]. Designing prompts for each contract type is costly and difficult to maintain [16, 17]. Ontological modeling, by contrast, enables reusable, adaptable extraction specifications across document types [12].

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan

*Corresponding author.

†These authors contributed equally.

kondor@g.ecc.u-tokyo.ac.jp (R. Kondo); shamik_kundu@trust-partner.co.jp (S. Kundu); muneaki_imai@trust-partner.co.jp (M. Imai); kyosuke_tomoda@trust-partner.co.jp (K. Tomoda); yuki-takazawa@g.ecc.u-tokyo.ac.jp (Y. Takazawa); hisanor@g.ecc.u-tokyo.ac.jp (R. Hisano)

https://trust-partner.co.jp/ (S. Kundu); https://trust-partner.co.jp/ (M. Imai); https://trust-partner.co.jp/ (K. Tomoda)

0009-0000-1190-6084 (R. Kondo); 0000-0002-5342-6262 (M. Imai); 0000-0002-4452-0302 (R. Hisano)

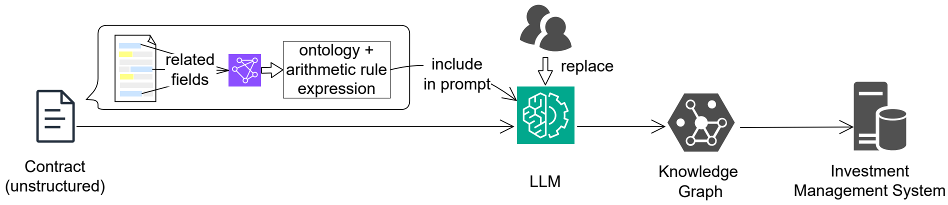

Figure 1: Overview of the information extraction workflow in financial operations.

Hence, this paper investigates how ontology design [18] and constraint specifications influence the performance of information extraction for financial contracts. Figure 1 provides an overview of our approach. We construct 12 prompt schemas for extracting information from financial documents, varying in representation style (natural language vs. ontology-based) and constraint inclusion (with vs. without rules). For the ontology-based variants, we extend Schema.org [19], aligning its vocabulary and structure with the Financial Industry Business Ontology (FIBO) [1]. In addition, we develop multiple constraint representations that vary in structural rigor and LLM-friendliness. We evaluate these prompt schemas using investment confirmation notices, which are highly structured and semantically rich documents, by embedding each schema into a prompt and measuring the LLM's extraction accuracy.

We make two key contributions. First, we show that incorporating ontologies and inter-property constraints into prompts improves LLM extraction performance, particularly for logically related properties. This contrasts with traditional approaches, where ontologies are typically used only for validation. Second, we find that less formal, LLM-adapted formats can outperform more structured representations, revealing a gap between conventional ontology design and the practical effectiveness of language models.

2. Methodology

Our approach is grounded in the hypothesis that the incorporation of ontological structures and inter-property constraints into prompts can enhance the accuracy and consistency of LLM-based information extraction. To evaluate this, we design a controlled experimental framework that systematically varies prompts used to guide extraction across two dimensions: representation style and constraint inclusion.

We use a curated corpus of 19 financial contract documents, consisting of 9 purchase confirmation notices and 10 sale confirmation notices[1]. Although these documents originate from custodial banking operations, they exhibit features that are often observed in financial contracts and notification documents more broadly, making them a useful sample for developing and assessing information extraction methods. These documents have the following notable characteristics. First, they vary considerably in format across firms, as there is no standardized template. Second, key information such as trade dates may be phrased in diverse ways. Third, such key information is often embedded within full sentences or tables and is intermingled with additional information. As a result, the documents are unstructured in nature, presenting diverse and realistic challenges for information extraction.

Each document contains 11 target properties. Of these, five are classified as *independent*(*Fund Code*, *Trade Date*, *Settlement Date*, *Base Currency*, and *Settlement Currency*), and six are classified as *dependent* properties that participate in arithmetic relationships:

$$\text{Order Quantity} \times \text{Unit Price} = \text{Gross Amount},$$
$$\text{Gross Amount} + \text{Fee} = \text{Settlement Amount (Settlement Currency)}.$$

Settlement Amount (Settlement Currency) represents the converted value of *Settlement Amount (Base Currency)* after currency exchange.

[1] Confirmation notices are widely used in custodial banking operations, where they are exchanged between securities firms and custodial banks as standard records of securities transactions. We edited the content while retaining key structural properties such as inter-field relationships.

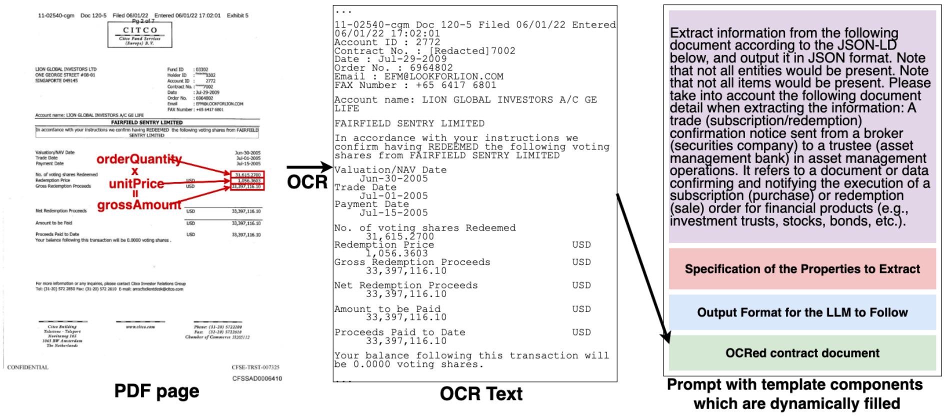

Figure 2: Overview from a contract PDF document to prompt with placeholders. Sample contract PDF is adapted from [20].

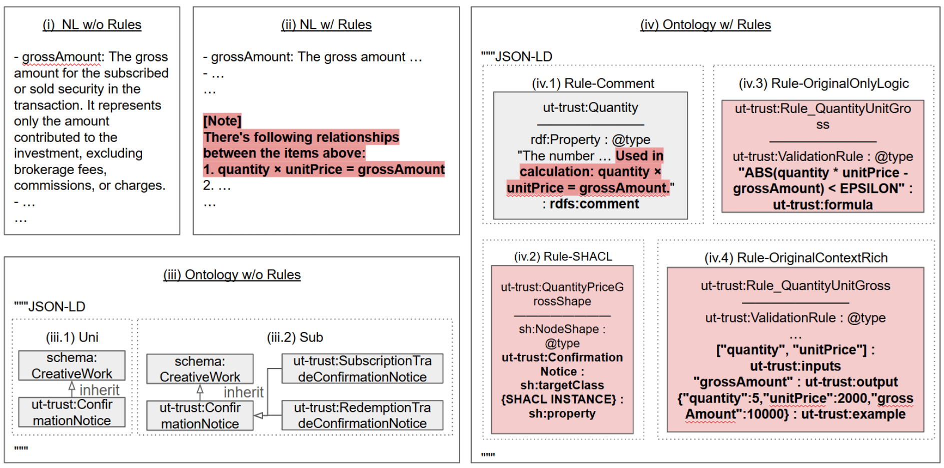

Figure 3: In the part of "Specification of the properties to extract", there are four schema categories. The red highlights show the constraints.

We construct prompts with four components: a task description, an OCRed contract, a list of properties to extract, and an output format (Figure 2). We vary only the property specification and output format across the twelve prompt schemas described below. The task description and document remain fixed. We parse LLM's responses accordingly and evaluate accuracy by exact match against the ground truth.

We create 12 prompt schemas varying along two dimensions: representation style (natural language or ontology-based), and inclusion of constraints (none or rules). This yields four schema categories: (i) natural language without rules, (ii) natural language with verbalized arithmetic constraints, (iii) ontology-based without constraints, and (iv) ontology-based with constraints (Figure 3). Within category (iii), we explore variations in ontological modeling across three aspects: class structure (separate classes for subscription/redemption — Subclassed vs. a unified class — Unified), transaction type representation (explicit property values — Tagged Type vs. natural language comments — Commented Type), and currency representation (enumerating allowed values — Enum Currency vs. comment-based guidance — Commented Currency). For category (iv), we evaluate four styles of constraint representation: natural language comments (Rule-Comment), formal SHACL rules (Rule-SHACL) [15], custom logic expressions (Rule-Only-Logic), and context-rich rules that include examples and guidance (Rule-Context-Rich).

These design variations allow us to evaluate how structural abstraction and constraint formalism influence LLM extraction performance, as well as to assess the sensitivity of outcomes to representational choices. Because small changes in prompts can lead to substantial shifts in LLM behavior, this diversity enables a robust evaluation of both the effectiveness and generality of different modeling strategies.

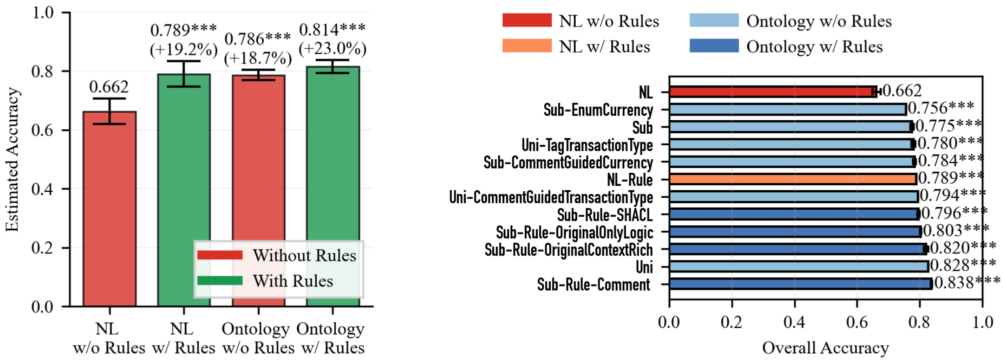

Figure 4: Performance comparison across schema categories and individual methods. (**Left**) Estimated overall accuracy by primary schema category. Linear mixed-effects model is used to estimate trial-level accuracies with category as a fixed effect (reference = NL w/o Rules) and method as a random intercept; error bars show 95% confidence intervals, stars denote significance versus the NL w/o Rules baseline (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). (**Right**) Method-level mean accuracy; error bars indicate ±2SD across trials and stars mark significant differences from the NL w/o Rules baseline. (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$) Method labels are abbreviated to make the display compact: **NL** = Natural Language, **S** = Subclassed, **U** = Unified.

We use the `gemma3n:e4b` model [21] with a 32k token context and a fixed temperature of 0.1. We ran each document–schema pair five times to account for randomness in the LLM output. We evaluated extraction accuracy using exact matching on each target property.

The source code used in our experiments is available at https://github.com/rkondo3/ut_trust_open.

3. Results

The left panel of Figure 4 shows the overall extraction accuracy for the four primary schema categories. Ontology-based representations outperformed natural language in both settings: without constraints, accuracy improved from 66.2% to 78.6% (+12.4%); and with constraints, from 78.9% to 81.4% (+2.5%). Similarly, adding rule-based constraints improved accuracy across representation styles: from 66.2% to 78.9% (+12.7%) for natural language, and from 78.6% to 81.4% (+2.8%) for ontology. The highest accuracy was achieved when both ontology and rules were included, which confirms their complementary effects.

The right panel of Figure 4 shows that all ontology + rule variants exceeded 80% accuracy, outperforming ontology-only and natural language baselines. Even the SHACL-based variant performs well, which suggests that structured knowledge improves extraction without manual tuning. Among these, the unstructured Rule-Comment schema achieved the highest accuracy (83.8%), whereas the structured SHACL variant was the lowest among the four (79.6%). This suggests that formal rigor does not always yield better LLM performance. With improved design, intermediate formats may offer a better balance between structure and model interpretability.

Our results also demonstrated that ontology structuring choices affected LLM performance. Unified-class schemas outperformed subclassed schemas, which aligns with Semantic Web practice that favors abstraction when distinctions are minimal [22]. Over-modeling may obscure patterns that LLMs could otherwise learn. For currency, comment-guided formats (e.g., "use ISO 4217 codes") outperformed formal definitions, which suggests that LLMs benefit from explicit prompts and motivates runtime embedding of linked ontology definitions.

Finally, the plain natural language schema without ontology or constraints performs worst (66.2%), confirming the substantial value of incorporating structural information—whether through ontological modeling, constraint specifications, or both.

Figure 5 provides a detailed breakdown of the property-level extraction accuracy. Independent properties such as *Fund Code*, *Base Currency*, and *Settlement Currency* were extracted with high accuracy across all schema categories (84% - 95%). Moreover, the average performance for the four approaches shows only moderate variations, ranging from 76% to 82% overall accuracy. These results suggest that

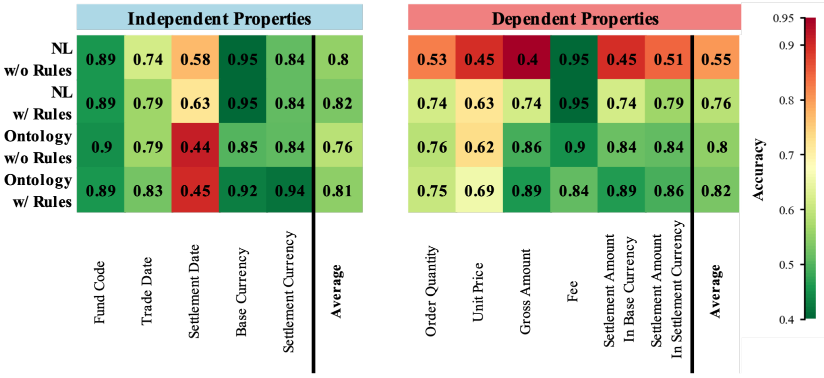

Figure 5: Field-level extraction accuracy across prompt schema types, separated by property type. The heatmap displays accuracy for each individual property (columns) under four prompt schema conditions (rows). Properties are grouped into **independent properties** (left) and **dependent properties** (right), based on whether their values can be inferred from other fields (e.g., arithmetic or logical dependencies).

for straightforward information retrieval, even simple natural language prompts are generally sufficient, and the added structure of ontologies or rules offers limited additional benefit.

Among the independent properties, date-related properties (*Trade Date*, *Settlement Date*) exhibited relatively higher difficulty. In particular, all approaches had low accuracy for *Settlement Date* (44% - 63%). When settlement date information was unavailable, these prompts frequently returned the trade date value instead of indicating missing information, thereby reducing extraction accuracy. This behavior suggests that, although ontological structure provides beneficial semantic guidance, it may also introduce systematic biases toward semantically related but distinct property values.

Dependent properties such as *Order Quantity*, *Unit Price*, *Gross Amount*, and *Settlement Amounts* proved relatively challenging for the NL w/o Rules approach with only 55% average accuracy. By contrast, the inclusion of either ontology or rules significantly improved results, raising accuracy to levels comparable to independent fields (76–82%). This underscores the importance of structural guidance in enabling LLMs to interpret relational constraints.

4. Conclusion

Using a diverse set of schema variants evaluated on financial documents, we demonstrated that prompts enriched with structural guidance through ontological modeling and inter-property constraints consistently outperformed purely natural language descriptions. This framework opens the possibility of quantitatively assessing ontological and constraint modeling choices through task-based evaluation.

5. Acknowledgments

R.H. is supported by JST FOREST Program (JPMJFR216Q), JST PRESTO Program (JPMJPR2469), Grant-in-Aid for Scientific Research (KAKENHI, JP24K03043) and the UTEC-UTokyo FSI Research Grant Program. R.K. is funded by JST, ACT-X Grant Number JPMJAX23CA, Japan. During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content. We also thank Edanz for editing the draft.

6. Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] Enterprise Data Management Council, Financial industry business ontology (FIBO), 2025.

[2] M. J. Schmidt-Kessen, H. Eenmaa, M. Mitre, Machines that make and keep promises - Lessons for contract automation from algorithmic trading on financial markets, Computer Law & Security Review 46 (2022) 105717.

[3] IBM, What is Finance Automation? | IBM, https://www.ibm.com/think/topics/finance-automation, 2025.

[4] T. Yamamoto, Practical handbook of generative AI in financial institutions, 2024.

[5] A. Satsiou, A. Revenko, I. Praggidis, E. Karapistoli, G. Panos, C. Bouzanis, I. Kompatsiaris, Semantic Technology for Financial Awareness, in: 12th International Conference on Semantic Systems (SEMANTiCS 2016), Leipzig, Germany, 2016.

[6] M. D. Flood, O. R. Goodenough, Contract as Automaton: The Computational Representation of Financial Agreements (2015).

[7] P. Casanovas, M. Palmirani, S. Peroni, T. van Engers, F. Vitali, P. Casanovas, M. Palmirani, S. Peroni, T. van Engers, F. Vitali, Semantic Web for the Legal Domain: The next step, Semant. web 7 (2016) 213–227.

[8] J. Breuker, P. Casanovas, M. C. Klein, E. Francesconi, Law, Ontologies and the Semantic Web – Channelling the Legal Information Flood, Frontiers in Artificial Intelligence and Applications, IOS Press, 2009.

[9] Y. He, J. Chen, H. Dong, I. Horrocks, Exploring large language models for ontology alignment, in: Posters & Demos of the 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, 2023. URL: https://arxiv.org/abs/2309.07172, iSWC 2023 Posters & Demos track.

[10] C. Saetia, J. Phruetthiset, T. Chalothorn, M. Lertsutthiwong, S. Taerungruang, P. Buabthong, Financial product ontology population with large language models, in: D. Ustalov, Y. Gao, A. Panchenko, E. Tutubalina, I. Nikishina, A. Ramesh, A. Sakhovskiy, R. Usbeck, G. Penn, M. Valentino (Eds.), Proceedings of TextGraphs-17: Graph-based Methods for Natural Language Processing, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 53–60. URL: https://aclanthology.org/2024.textgraphs-1.4/.

[11] Trust Co., Ltd., Trust GenGA, https://trust-partner.co.jp/TrustGenGA, 2025. Accessed: 2025-07-31.

[12] M. Bennett, The financial industry business ontology: Best practice for big data, Journal of Banking Regulation 14 (2013) 255–268.

[13] B. Motik, P. F. Patel-Schneider, B. Parsia, OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition), W3C Recommendation, 2012. URL: https://www.w3.org/TR/owl2-syntax/, available at: https://www.w3.org/TR/owl2-syntax/.

[14] A. Omran, V. Janev, A. Fensel, S. Auer, Learning shacl shapes from knowledge graphs, Semantic Web 13 (2022) 877–907. URL: https://doi.org/10.3233/SW-210431. doi:10.3233/SW-210431.

[15] W3C RDF Data Shapes Working Group, Shapes Constraint Language (SHACL), W3C Recommendation, 2017. URL: https://www.w3.org/TR/shacl/, available at https://www.w3.org/TR/shacl/.

[16] J. Zhuo, S. Zhang, X. Fang, H. Duan, D. Lin, K. Chen, Prosa: Assessing and Understanding the prompt sensitivity of LLMs, in: Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 1950–1976. URL: https://aclanthology.org/2024.findings-emnlp.108/. doi:10.18653/v1/2024.findings-emnlp.108.

[17] A. Chatterjee, H. K. Renduchintala, S. Bhatia, T. Chakraborty, Posix: A prompt sensitivity index for large language models, in: Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, 2024, pp. 14550–14565.

[18] A. Gangemi, Ontology design patterns for semantic web content, in: Proceedings of the 4th International Semantic Web Conference (ISWC 2005), volume 3729 of *Lecture Notes in Computer Science*, Springer, 2005, pp. 262–276. doi:10.1007/11574620_21.

[19] Schema.org Community, Schema.org, https://schema.org/, 2011. Accessed: 2025-07-31.

[20] Exhibit 5: Lion global investors ltd redemption statement, 11-02540_cgm doc 120-

5, https://www.madofftrustee.com/images/upload/11-02540_Lion_Global_Investors_Beckerlegge_Declaration_Exh5_120-0005.pdf, 2022. Accessed 2025-09-01.

[21] Gemma Team, Gemma 3n: Model documentation, 2025.

[22] N. F. Noy, D. L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, Technical Report KSL-01-05 & SMI-2001-0880, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, 2001. URL: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf, technical Report.