

# LLM を用いた不動産情報抽出における ドメイン知識の段階的付与が精度に与える影響

松下 莉久<sup>1,2</sup> 本郷 慎一<sup>2</sup> 叶内 晨<sup>2,3</sup> 岩成 達哉<sup>2,4</sup>  
<sup>1</sup> 東洋大学 <sup>2</sup> 株式会社 estie <sup>3</sup> NLPeanuts Inc. <sup>4</sup> 東京大学  
s1f102300571@iniad.org, shin.kanouchi@nlpeanuts.com  
{shinichi.hongo, tatsuya.iwanari}@estie.co.jp

## 概要

本研究では、不動産業界における文書からの情報抽出において、大規模言語モデル (LLM) に対するドメイン知識の段階的付与が精度に与える影響を調査する。不動産分野のような専門性の高い領域では適切なドメイン知識の提供が重要だが、どの程度・どのような知識を LLM に与えるべきかは十分に研究されていない。本研究では、J-REIT のプレスリリース PDF を対象とし、土地面積・建蔽率・容積率を抽出する3つのタスクにおいて、知識を専門性に応じて段階的に付与し精度を検証した。実験の結果、用語の関係性定義やドメインエキスパート知識などの専門性の高い知識の LLM への付与が、抽出精度向上に大きく寄与することを示した。

## 1 はじめに

不動産業界では、物件情報や取引データなどが PDF ファイルで流通しており、これらの情報を効率的に構造化することは業務効率化の観点から重要である。その中でも、不動産投資法人 (J-REIT) に関連する取引や鑑定情報を含むプレスリリース資料は、数少ない公開取引事例として注目されている。

近年、大規模言語モデル (LLM) を用いた非構造化データからの情報抽出が注目されており、プロンプトにドメイン知識を含めることで性能が向上することが知られている [1]。不動産業界でも LLM を用いた情報抽出の研究が進められている [2, 3] が、一方で、どの程度・どのような種類の知識を提供すべきかについては十分に体系化されていない。特に不動産特有の用語の定義、計算方法、業界慣習など多様な知識が存在するが、これらを段階的に付与した際の効果は定量的に検証されていない。

本研究では、LLM に対してドメイン知識を段階

的に付与することで情報抽出精度がどのように変化するかを体系的に検証する (図 1)。具体的には、最小限の知識からドメインエキスパート知識まで4段階のレイヤーを設計し、J-REIT のプレスリリース PDF に対する3つの情報抽出タスク (土地面積、建蔽率、容積率) で評価する。これらのタスクは、複雑な計算や注釈の読み取りなど異なる特性を持ち、知識付与の効果を多面的に分析できる。

実験の結果、LLM に対する用語定義の付与では精度向上が見られないが、関係性定義とドメインエキスパート知識の付与で大幅に向上し、全タスクで88%以上の精度を達成した。この結果から、LLM は不動産用語を事前学習で獲得しているが、関係性定義やドメインエキスパート知識は獲得しておらず、明示的な知識付与が必要なことが示された。さらに、土地面積タスクでは知識付与により推論トークン数が大幅に増加する一方、建蔽率・容積率タスクでは増加が限定的であり、タスクの特性によって推論プロセスへの影響が異なることを示した。

## 2 関連研究

大規模言語モデル (LLM) は自然言語処理の分野で顕著な進展を遂げており [4, 5, 6]、非構造化テキストの情報抽出や構造化タスクにおいて高い性能を発揮している [7]。近年では画像や PDF ファイルを直接入力可能なマルチモーダル LLM も登場しており [8]、より柔軟なデータ入力が可能となっている。

LLM の性能を最大化するプロンプトエンジニアリングは重要な研究領域となっており [9]、Chain-of-Thought (CoT) プロンプティング [10] や Few-shot learning [11] などの手法が提案されている。ドメイン固有の知識をプロンプトに統合する研究も進められており、化学ドメインではドメイン知識を埋め込んだプロンプトエンジニアリングが従来の戦

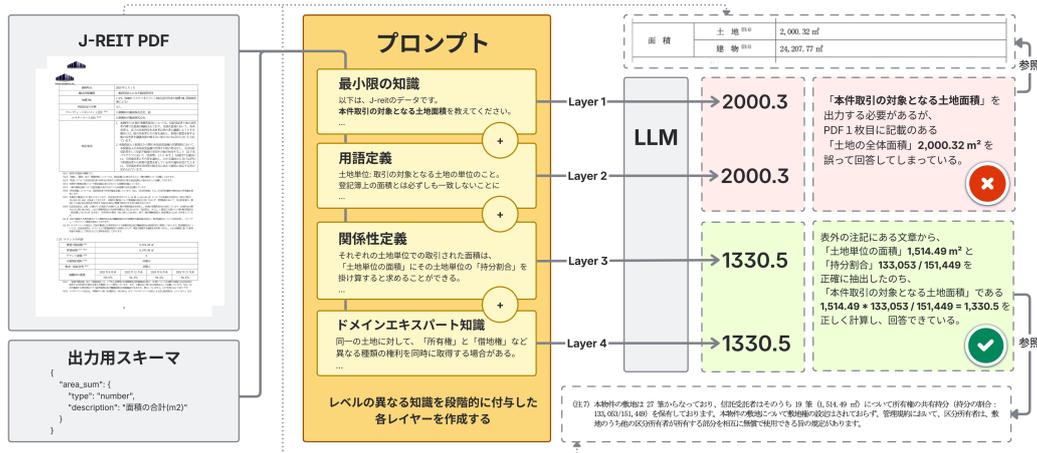


図1 本研究の全体像（土地面積タスクの例）

略を上回ることが実証されている [12]. しかし、これらの研究ではドメイン知識を一度にすべて提供するか、モデル自身が知識を生成するアプローチが主流であり、知識を段階的に付与した際の効果を体系的に分析した研究は限られている。

不動産業界では、LLMを用いた情報抽出の研究が進められている。Kvetら [2] は不動産広告からの属性抽出において、属性定義や文選択などの入力設計により高い抽出性能を実現した。Zhaoら [3] は不動産契約からの自動情報抽出にLLMを適用し、ファインチューニングによる性能向上を示した。また日本では、マイソクPDFからの情報抽出 [13, 14] や複数LLMのアンサンブル手法による不動産情報抽出業務の効率化 [15] などが報告されている。本研究は、知識を段階的に付与し、各段階での精度を定量評価する点で新規性がある。

### 3 タスク概要

本研究では、J-REITのプレスリリースPDFからの情報抽出タスクに取り組む。J-REITプレスリリースは、数少ない公開取引事例として不動産業界で重要視されているが、多様な書式・レイアウト、専門用語・略語の混在、複雑な数値情報、注釈問題など、自動抽出が困難な特徴を持つ。

LLMによる推論時の知識付与の効果を多面的に検証するため、性質の異なる3つのタスクを設計した。土地面積タスクは、取引対象となる土地の総取引面積を取得するタスクであり、権利の種類や持分割合に応じた計算を要する。建蔽率タスクおよび容積率タスクは、用途地域と実質建蔽率・実質容積率のペアを抽出するタスクで、注釈の読み取りや加重平均の理解を要する。

## 4 提案手法

本研究では、ドメイン知識を4段階に体系化し、各段階におけるLLMの情報抽出性能を評価する。ドメイン知識は性質の異なる複雑な情報から構成されており、用語の定義のような基礎的知識から、それらの用語間の関係性、さらには対象文書の読み解き方に関する経験的・実務的知識まで多岐にわたる。本研究では、これらの知識を段階的に整理し、以下の4層に分類した：

1. 最小限の知識：タスク概要と出力スキーマのみ
2. 用語定義：タスク固有の用語の説明
3. 関係性定義：用語間の関係性、計算式
4. ドメインエキスパート知識：対象文書の読み解き方、実務上のノウハウ

各知識層は下位層を前提として累積的に適用し、4層目ではすべての知識が統合された状態となる。これにより、どの段階の知識が精度向上に寄与するかを定量的に評価できる。

### 4.1 Layer 1：最小限の知識

タスク概要と出力スキーマのみを提供し、ドメイン知識は付与しない。LLMが事前に獲得している知識のみでどの程度タスクを解けるかを確認する。なお自動評価のため、データ型や単位などを含む出力のために必要な情報は、出力用JSONスキーマとして全層で共通してLLMに付与する。

プロンプト例（土地面積）：「以下は、J-REITのデータである。本件取引の対象となる土地面積を教えてください。」

## 4.2 Layer 2：用語定義

Layer 1に加えて、タスクに必要な用語の定義を追加する。各用語について、その意味と役割を1~2文で説明している。用語の例を以下に示す。

- **土地面積タスク**：土地単位、保有形態、信託受益権、権利割合、持分割合
- **建蔽率・容積率タスク**：用途地域、実質建蔽率、指定建蔽率、実質容積率、指定容積率、基準容積率

## 4.3 Layer 3：関係性定義

Layer 2に加えて、用語間の一般的な関係性や計算式を追加する。関係性の例を以下に示す。

### 土地面積タスク：

- 土地単位ごとに「権利割合」と「保有形態」が存在
- 保有形態が「信託財産」の場合のみ「信託受益権割合」が存在
- 持分割合 = 権利割合 × 信託受益権割合（未定義の場合は1）
- 取引面積 = 土地単位の面積 × 持分割合
- 複数の土地単位がある場合は合算

### 建蔽率・容積率タスク：

- 用途地域は敷地に紐づく属性であり、指定建蔽率・指定容積率の前提となる
- 実質建蔽率 = 指定建蔽率 + 緩和条件による加算
- 実質容積率 =  $\min(\text{指定容積率}, \text{基準容積率})$
- 複数用途地域にまたがる場合は加重平均を適用

## 4.4 Layer 4：ドメインエキスパート知識

Layer 3に加えて、対象文書特有の処理ルールや注釈の読み方など、ドメインエキスパートが実務で培った知識を追加する。これらは不動産領域の一般的なドメイン知識に限らず、不動産文書を扱う際の実践的なノウハウを含む。以下に例を示す。

### 土地面積タスク：

- 数値が「計算前（地積）」なのか「計算後（取得面積）」なのかは、項目名などから判断する
- 同一土地に異なる権利を同時に取得する場合は、それぞれ独立して計算し合算する

### 建蔽率・容積率タスク：

- 注釈（※、注1）などに緩和条件や加重平均

値が記載されている場合がある

- 複数用途地域の表記は「商業地域／近隣商業地域」や「(A) ○○% (B) ○○%」など多様
- 「80/400」のように建蔽率/容積率をスラッシュでまとめて記載する場合がある

## 5 実験

### 5.1 実験設定

実験には、J-REITのプレスリリースPDFを使用した。各タスクについて30件の文書を用意し、人手で正解を付与した。このうち10件をプロンプトチューニング用、20件を評価用とした。

使用モデルはGemini 2.5 Pro[6]（パラメータはデフォルト設定）であり、マルチモーダル入力機能を利用した。各層のプロンプトとPDFと出力スキーマをモデルへ入力し、JSON形式で結果を得た。土地面積タスクでは総取引面積（m<sup>2</sup>）を数値で、建蔽率・容積率タスクでは用途地域名と値の組を配列で出力した。

精度の評価のは、完全一致による正解率（Accuracy）を用いた。ただし、数値情報は誤差を考慮し、有効数字3桁に丸めた値で一致判定を行った。また、LLMの出力のばらつきを考慮し、各条件で3回推論を行い、その平均値から正解率およびトークン数を算出した。

### 5.2 実験結果

表1に各タスクの実験結果を示す。

**Layer 1-2** 土地面積タスクでは、Layer 1（31.7%）とLayer 2（30.0%）で精度がほぼ変わらず低い結果となった。これは、用語の定義のみでは計算や推論を正しく行うための情報として不十分であり、計算式や処理ルールの付与が重要であることを示している。一方、建蔽率・容積率タスクではLayer 1で既に高精度（86.7%/80.0%）を達成しており、LLMの事前知識で対応可能であった。

また、全タスクでLayer 2ではLayer 1よりも推論トークン数が増加したが、精度はいずれも低下した。今回付与した用語定義は、モデルにとって既知の情報であり、追加の効果がなかっただけでなくプロンプトの複雑化により推論を阻害した可能性がある。

**Layer 2-3** 土地面積タスクでは、Layer 3において精度が80.0%へと大幅に向上した。これは、持分

表1 タスク毎の正解率 (Accuracy) とトークン数

タスク	指標	Layer 1	Layer 2	Layer 3	Layer 4
土地面積	正解率	31.7%	30.0%	80.0%	88.3%
	入力トークン数	696	959	1,347	1,700
	推論トークン数	788	921	2,783	2,581
建蔽率	正解率	86.7%	85.0%	90.0%	100.0%
	入力トークン数	708	785	953	1,179
	推論トークン数	691	819	838	850
容積率	正解率	80.0%	75.0%	88.3%	90.0%
	入力トークン数	706	808	968	1,192
	推論トークン数	801	1,060	1,041	952

割合の計算方法を明示したことで、単純な値の抽出では対応できなかったケースが正解となったためである。同様に、容積率タスクでも、Layer 3で75.0%から88.3%へと向上した。これは「指定容積率と基準容積率の小さい方を採用する」という条件（関係性）の提示が効果的であったと考えられる。一方で、Layer 3においても不正解となった事例は、対応ルールを網羅的に形式化することが困難な特殊ケースであった。以上の結果から、導出方法が一般的な場合は、Layer 3相当の知識を付与することで、高精度に情報抽出可能なことが示された。

**Layer 3-4** Layer 4では、全タスクで高い精度を達成した（土地面積88.3%、建蔽率100.0%、容積率90.0%）。これは、注釈の読み取りや加重平均の処理などといった、ドメインエキスパート知識の付与が効果的であったためと考えられる。

**推論トークン数の変化** 推論トークン数はタスクによって異なる傾向を示した。計算が主体となる土地面積タスクでは、推論トークン数がLayer 1-2 (788, 921) からLayer 3-4 (2,783, 2,581) で約3倍に増加した。関係性の付与により、モデルが計算過程を推論するようになったと考えられる。ただし、正解事例と不正解事例のトークン数を事例毎に確認すると、正解事例は比較的少ない推論トークン数で解に至る一方、不正解事例はトークン数が多い傾向があった。例えばLayer 4では、正解事例は2000トークン前後に対し、不正解事例は4000トークン前後であった。この傾向はLayerが深くなるほど顕著で、難しい事例でモデルが解を探索し続け、推論が長引く様子が観察された。抽出が主体となる建蔽率・容積率タスクでは、推論トークン数はLayer 2で増加した後、Layer 4までほぼ変わらなかった。

### 5.3 タスク毎のエラー分析

**土地面積タスク** Layer 1-2で正解した事例は、正解の土地面積が直接文書に記載されている場合に

限られ、それ以外では抽出誤りや持分割合の掛け忘れが多発した。Layer 3で関係性定義を与えることでこれらのミスは解消し精度が大幅に向上したが、複雑な計算や参照が求められる場合は依然不正解となった。Layer 4にて、複数の権利（所有権と敷地権など）の合算が必要な場合などのドメインエキスパート知識を付加することで改善したが、特殊ケースは複雑ですべて正解するには至らなかった。

**建蔽率タスク** 本文中や注釈に正解となる建蔽率が直接記載されている場合が多く、Layer 1から高い精度を達成した。主なエラーは、防火地域や準防火地域を用途地域と誤認識する場合と、複数の用途地域を勝手に結合して出力する場合であった。これらはLayer 3の関係性説明で大部分が改善され、Layer 4で完全に解消された。

**容積率タスク** 建蔽率タスクと同様に正解が直接記載される場合が多く、Layer 1からある程度高い精度を達成した。エラーパターンとしては、指定容積率と基準容積率の誤選択、加重平均の誤計算、防火地域の誤認識などがあった。誤認識や誤計算はLayer 3で解消されたが、同一用途地域内に複数の容積率が存在するケースなど、Layer 4でも一部のエラーが残存した。

## 6 結論

本研究では、J-REIT プレスリリースからの情報抽出タスクにおいて、ドメイン知識を4段階で付与し、各段階での精度向上を検証した。その結果、用語定義だけでは不十分であり、用語の関係性定義やドメインエキスパート知識のLLMへの明示的な付与が精度向上に大きく寄与することを示した。また、タスクごとに必要な知識の種類や深さが異なることが明らかになり、知識付与の設計にはタスク特性の考慮が重要であることが示唆された。今後は、マイソクや契約書など他の不動産文書への適用を通じて、知見の汎用性を検証していきたい。

## 参考文献

- [1] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, et al. A comprehensive survey on pre-trained foundation models: A history from bert to chatgpt. **International Journal of Machine Learning and Cybernetics**, pp. 1–65, 2024.
- [2] Michal Kvet, Miroslav Potočár, and Slavomír Tatarka. Real estate attribute value extraction using large language models. **IEEE Access**, Vol. 13, pp. 5816–5831, 2025.
- [3] Yu Zhao, et al. Utilizing large language models for information extraction from real estate transactions. **arXiv preprint arXiv:2404.18043**, 2024.
- [4] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. **arXiv:2303.08774**, 2023.
- [5] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. **arXiv preprint arXiv:2302.13971**, 2023.
- [6] Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. **arXiv:2507.06261**, 2025.
- [7] Alexander Dunn, John Dagdelen, Nicholas Walker, Sanghoon Lee, Andrew S. Rosen, Gerbrand Ceder, Kristin Persson, and Anubhav Jain. Structured information extraction from complex scientific text with fine-tuned large language models. **arXiv:2212.05238**, 2022.
- [8] Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. Gemini: a family of highly capable multimodal models. **arXiv:2312.11805**, 2023.
- [9] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. **arXiv:2102.07350**, 2021.
- [10] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. **Advances in neural information processing systems**, Vol. 35, pp. 24824–24837, 2022.
- [11] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. **Advances in neural information processing systems**, Vol. 33, pp. 1877–1901, 2020.
- [12] Hongxuan Liu, Haoyu Yin, Zhiyao Luo, and Xiaonan Wang. Integrating chemistry knowledge in large language models via prompt engineering. **Synthetic and Systems Biotechnology**, Vol. 10, No. 1, pp. 23–38, 2024.
- [13] 本郷慎一, 叶内晨, 齊藤佑太郎, 岩成達哉. 大規模言語モデルを用いたマイソク PDF からの情報抽出. 言語処理学会第 30 回年次大会予稿集 (NLP2024), 2024.
- [14] 樋口莉央, 叶内晨, 齊藤佑太郎, 松本健太郎, 岩成達哉. 大規模言語モデルを用いたマイソク PDF からの情報抽出の改善. 人工知能学会全国大会予稿集 (JSAI2025), 2025.
- [15] 齊藤佑太郎, 叶内晨, 松本健太郎, 岩成達哉. 不動産情報抽出業務の効率化に向けた大規模言語モデルを用いたアンサンブル手法. 言語処理学会第 31 回年次大会予稿集 (NLP2025), 2025.