Theme-Explanation Structure for Table Summarization using Large Language Models: A Case Study on Korean Tabular Data

TaeYoon Kwack[1*] **Jisoo Kim**[1*] **Ki Yong Jung**[1] **DongGeon Lee**[2] **Heesun Park**[1†]
[1]Sungkyunkwan University [2]Pohang University of Science and Technology
{njj05043, clrdln, wjdrldyd0213}@g.skku.edu
donggeonlee@postech.ac.kr hspark20@skku.edu

Abstract

Tables are a primary medium for conveying critical information in administrative domains, yet their complexity hinders utilization by Large Language Models (LLMs). This paper introduces the **T**heme-E**x**planation Structure-based Table Summarization (**Tabular-TX**) pipeline, a novel approach designed to generate highly interpretable summaries from tabular data, with a specific focus on Korean administrative documents. Current table summarization methods often neglect the crucial aspect of human-friendly output. Tabular-TX addresses this by first employing a multi-step reasoning process to ensure deep table comprehension by LLMs, followed by a journalist persona prompting strategy for clear sentence generation. Crucially, it then structures the output into a Theme Part (an adverbial phrase) and an Explanation Part (a predicative clause), significantly enhancing readability. Our approach leverages in-context learning, obviating the need for extensive fine-tuning and associated labeled data or computational resources. Experimental results show that Tabular-TX effectively processes complex table structures and metadata, offering a robust and efficient solution for generating human-centric table summaries, especially in low-resource scenarios.

1 Introduction

Tables are essential for presenting core information, especially within the administrative domain, where critical data is frequently structured in tabular formats (Musumeci et al., 2024). The ability of Large Language Models (LLMs) to accurately summarize and elucidate the contents of these tables is becoming increasingly significant for data utilization. An important aspect of effective table summarization is the generation of human-understandable output. This necessitates not only the LLM's profound comprehension of the input table, but also the crafting of summaries that are both intuitive and concise, delivering key information without ambiguity.

Despite the critical need for human-centric summaries, recent research in table-to-text generation has often overlooked this aspect. Many existing approaches prioritize other metrics or model architectures, without sufficiently addressing how the generated text will be perceived and understood (Liu et al., 2024; Zhang et al., 2024b). Consequently, while the output can be factually correct, they lack the clarity, conciseness, or intuitive structure that facilitates effortless human comprehension, particularly when dealing with specialized data such as Korean administrative tables (NIKL, 2024).

To address this gap and emphasize the generation of human-friendly summaries, we propose the **T**heme-E**x**planation Structure-based Table Summarization (**Tabular-TX**) pipeline. Our approach is meticulously designed to guide LLMs towards producing summaries that are not only accurate but also exceptionally interpretable. First, to ensure a deep understanding of the input table, we decompose the LLM's reasoning process into a multi-step procedure, where each step focuses on a specific inferential task, thereby simplifying the complex table interpretation process. Second, we employ a "Journalist Persona" prompting strategy to encourage the generation of clear, objective, and well-phrased sentences. Finally, and most critically, we transform these generated insights into a highly structured Theme-Explanation (TX) format, where each summary segment consists of a thematic adverbial phrase followed by a predicative explanatory clause, enhancing readability and directness.

The Tabular-TX pipeline offers significant advantages, particularly in resource-constrained environments where extensive fine-tuning is unfeasible. Our primary contributions are threefold:

- We introduce a novel pipeline that leverages

[*]Both authors contributed equally to this work.
[†]Corresponding author.

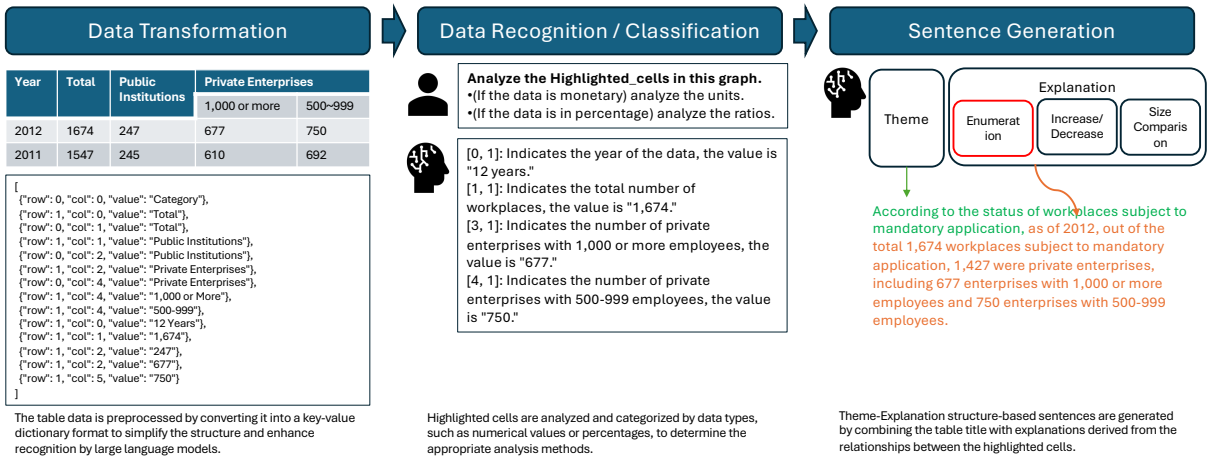

Figure 1: An overall pipeline of Theme-Explanation Structure-based Table Summarization (Tabular-TX).

multi-step reasoning and a journalist persona to generate high-quality textual explanations from complex tabular data.

- We propose the Theme-Explanation (TX) sentence structure, a new format for table summaries designed to maximize human interpretability.
- Through empirical evaluation, we demonstrate that our In-Context Learning (ICL)-based approach enables LLMs to achieve strong performance in table data processing and summarization without the need for task-specific fine-tuning.

2 Related Work

Table-to-Text Generation To enable complex reasoning over tabular data, Wang et al. (2024) proposed the Chain-of-Table framework as an extension of the text-based Chain-of-Thought method (Wei et al., 2022). This approach simplifies inference by reordering, extracting, and filtering table data, ultimately integrating relevant information into a structured table format. While this method excels in structured table processing and mathematical reasoning, it has limitations in generating interpretations for sections that require metadata or background knowledge.

TableLlama (Zhang et al., 2024b) aims to generalize table-based models beyond task-specific constraints by fine-tuning 14 datasets across 11 tasks, including Highlighted Cells question-answering. The model achieved performance comparable to or surpassing task-specific models and even outperformed GPT-4 (OpenAI, 2023) on unseen tasks. However, despite its effectiveness, the model suffers from high computational costs.

Among table-based interpretation benchmarks, FeTaQA (Nan et al., 2022) serves as a key reference dataset. While Chain-of-Table and TableLlama utilize in-context learning (Brown et al., 2020) and fine-tuning, respectively, they struggle to incorporate metadata into their interpretations effectively.

3 Theme-Explanation Structure

Unlike conventional approaches that treat table summaries as isolated text generation tasks, our method ensures structural consistency by explicitly organizing content into a **Theme Part** and an **Explanation Part**.

3.1 Theme Part

The Theme Part serves as a crucial contextual anchor, ensuring that numerical or categorical values in the table are interpreted correctly. It is structured as an adverbial phrase, combining the noun phrase of the table title (`table_title`) with a citation or basis expression[1]. This structure is essential because the table title provides the sole comprehensive context in table summaries. Unlike general text summaries, table cells alone are insufficient to provide meaningful context, making the resulting sentence ambiguous without additional background information.

3.2 Explanation Part

Following the Theme Part, the Explanation Part delivers a structured analysis of the highlighted cells,

[1]Comparison of summarization sentence with and without theme part is illustrated at Figure 4.

forming the core content of the summary. Depending on the data type, this section uses a specific analytical technique, such as enumeration, magnitude comparison, or trend analysis. The choice of method is determined based on the comparability of the highlighted cells, ensuring that the summary provides meaningful insights rather than just raw cell values.

4 Tabular-TX Pipeline

Generating summaries with the theme-explanation structure in Tabular-TX involves multiple processing steps to transform tabular data into structured natural language summaries.[2]

4.1 Chain-of-Thought (CoT)

After data transformation, the actual sentence generation process begins using Chain-of-Thought (CoT) reasoning. Large Language Models (LLMs) generally face performance degradation when handling tabular data summarization, as this task simultaneously requires multiple capabilities such as table recognition, mathematical reasoning, and commonsense inference. This issue, known as the Compositional Deficiency problem (Zhao et al., 2024), occurs because individual data points tend to be analyzed separately without adequately integrating their relationships into a holistic interpretation. CoT mitigates this by systematically guiding the model to tackle one reasoning step at a time, thereby improving interpretative accuracy and contextual completeness.

Specifically, CoT first classifies the types of highlighted cells, distinguishing among monetary values, percentages, categorical data, and textual explanations. This classification step prevents potential errors, such as misinterpreting percentages as plain numbers. Then, depending on the classified data type, the most appropriate analytical method—such as enumeration for listing individual items, magnitude comparison for numerical rankings, or trend analysis for temporal changes—is selected and applied. For instance, monetary values are converted into consistent units, and percentages are appropriately formatted for clarity.

In the context of Korean administrative table data, these challenges are further complicated by the language's implicit nature, the potential gap between administrative and everyday terminology, and morphological complexities (e.g., absent subjects or ambiguous particle usage). CoT systematically decomposes these linguistic hurdles by (1) classifying specialized terms, (2) normalizing numeric expressions in line with Korean usage conventions, and (3) incrementally integrating contextual cues, such as clarifying administrative vocabulary or disambiguating omitted referents. Through this stepwise process, the model avoids misinterpretations caused by either unfamiliar terms or implicit structures, ultimately generating summaries that better align with Korean textual norms.

By addressing each reasoning subtask explicitly and sequentially, CoT ensures the final table summary captures data relationships clearly and coherently, resulting in an accurate and contextually meaningful summary.

4.2 Journalist Persona for Structured Generation

In last step of pipeline we assign a journalist persona to the LLM to generate Theme-Explanation structured summaries. This persona is particularly effective because table summaries share key characteristics with straight news articles, which prioritize conciseness, objectivity, and fact-based clarity. Rather than generating overly detailed or speculative content, the model produces well-structured and neutral summaries that adhere to journalistic reporting conventions when guided by this persona.

Figure 2 demonstrates the impact of the journalist persona on table summarization. With a generic prompt, the model generates an ambiguous summary that captures core information but lacks contextual clarity and coherence. In contrast, applying the journalist persona produces a structured and contextually enriched summary. This improvement occurs because the journalist persona explicitly guides the model to state information sources, clearly define numerical constraints, and incorporate contextual details, closely resembling news article formats.

5 Experiment

5.1 Experimental Setup

Dataset For training and evaluation, we utilized the Korean table interpretation benchmark (NIKL, 2024), which focuses on summarizing highlighted table segments into coherent sentences. An example of the dataset is shown in Figure 3. The dataset consists of 7,170 training tables, 876 validation ta-

[2]Within Tabular-TX, data is preprocessed before using LLMs. Further details on data preprocessing are discussed in Appendix C.1.

Model	ROUGE-1	ROUGE-L	BLEU	Average
KoBART - Fine-tuned	0.37	0.28	0.35	0.33
EXAONE 3.0 7.8B - ICL	0.21	0.14	0.01	0.12
EXAONE 3.0 7.8B - LoRA	0.27	0.21	0.05	0.17
EXAONE 3.0 7.8B - Tabular-TX	**0.51**	**0.39**	**0.44**	**0.45**
llama-3-Korean-Bllossom-8B - ICL	0.33	0.25	0.27	0.28
llama-3-Korean-Bllossom-8B - Tabular-TX	**0.48**	**0.37**	**0.42**	**0.43**

Table 1: Evaluation results on the Korean Korean table interpretation benchmark for each model.

bles, and 876 test tables. Each data point contains metadata such as the document title, table title, publication date, publishing organization, table source URL, highlighted cell information, table data, and a reference summary sentence describing the highlighted portions' key contents.

Evaluation Metrics We employ ROUGE-1, ROUGE-L, and BLEU to evaluate the performance of table segment interpretation. These metrics assess how effectively the summaries convey the key content of the table while achieving high semantic quality.

Models To evaluate the effectiveness of the Tabular-TX pipeline, we utilize EXAONE 3.0 7.8B (An et al., 2024) and llama-3-Korean-Bllossom-8B[3] models as base models. EXAONE 3.0 7.8B, a successor to EXAONE-LM-v1.0, has demonstrated state-of-the-art performance in Korean TableQA, ranking first on the KorWikiTableQuestions (Jun et al., 2022). Similarly, llama-3-Korean-Bllossom-8B is the top-performing sub-10B model in a Korean multi-domain reasoning benchmark. We compare the performance of these models with and without Tabular-TX, assessing whether structured generation enhances performance in table summarization. Additionally, we analyze whether Tabular-TX reduces reliance on extensive fine-tuning while maintaining high-quality summaries.

5.1.1 Additional Adaptation Approaches

In-Context Learning (ICL) We begin by applying ICL to each model, providing a few table-summarization examples without explicit fine-tuning. This approach tests how effectively the model can generate coherent sentences for highlighted table cells based solely on a small set of demonstrations.

Low-Rank Adaptation (LoRA) Next, we assess the computational efficiency and performance of the Tabular-TX pipeline by introducing LoRA (Hu et al., 2022). We trained the EXAONE 3.0 7.8B by applying LoRA to see if we could maintain high-quality table summaries with fewer resources.

Full Model Fine-Tuning Finally, we evaluate the KoBART (Korean BART)[4] model under a full model fine-tuning setup to determine whether a smaller-scale language model can achieve comparable performance when all its parameters are trained on the Korean table interpretation benchmark.

5.2 Experimental Results

Table 1 presents the performance of various models evaluated using ROUGE-1, ROUGE-L, BLEU, and their average scores. The KoBART recorded an average score of 0.33 after fine-tuning. In contrast, EXAONE 3.0 7.8B achieved 0.12 with the ICL method, 0.17 after fine-tuning, and 0.45 when combined with the Tabular-TX method. Similarly, llama-3-Korean-Bllossom-8B, which was also tested with Tabular-TX, showed a notable improvement, reaching an average score of 0.43. These results demonstrate that Tabular-TX consistently outperforms alternative methods, achieving the highest overall performance across different model configurations.

The performance gap between EXAONE 3.0 7.8B and KoBART, despite both being fine-tuned on the same dataset, can be explained through the multiplicative joint scaling law (Zhang et al., 2024a). This principle suggests that when the dataset size is insufficient relative to the model size, the performance gains from fine-tuning remain limited. Since KoBART has 124M parameters, while EXAONE 3.0 is approximately 63 times larger, the

[3] https://huggingface.co/MLP-KTLim/llama-3-Korean-Bllossom-8B

[4] https://huggingface.co/gogamza/kobart-base-v2

dataset required to achieve a comparable performance boost must be proportionally scaled up by a factor of 63. The inability to meet this scaling requirement explains why the performance of the fine-tuned KoBART plateaued, while EXAONE 3.0 7.8B demonstrated more significant gains with the same dataset.

This study confirms that the proposed Tabular-TX method enhances table data analysis performance without fine-tuning. Notably, Tabular-TX outperforms traditional fine-tuned models despite relying on significantly smaller datasets, demonstrating its efficacy in resource-constrained learning environments. Moreover, Tabular-TX achieved approximately four times higher average performance compared to standard ICL methods, further reinforcing its role as a scalable and efficient alternative for structured table summarization tasks.

6 Conclusion

This study introduced the Theme-Explanation Table Summarization (Tabular-TX) pipeline, a novel approach to improve table summarization tasks with low-resource requirements. Moreover, the proposed pipeline effectively overcame unique challenges in Korean administrative table processing.

Experimental results signify that Tabular-TX enhances table summarization performance. This study contributed to the summarization of complex table data by introducing a novel sentence generation method based on the theme-explanation structure. Furthermore, Tabular-TX achieved excellent performance without fine-tuning, by incorporating ICL. This indicates its potential as a significant contribution to table data analysis, even in resource-constrained environments, without requiring direct model training.

Limitations

We acknowledge a few limitations in this study. First, Tabular-TX was only evaluated on EXAONE 3.0 7.8B and llama-3-Korean-Bllossom-8B, leaving the question of its effectiveness across a broader range of LLMs open. Second, this study primarily focused on Korean administrative table data, and further research should investigate whether the Theme-Explanation structure is equally effective for diverse tabular data formats in other languages or specialized domains. Finally, Tabular-TX currently relies on predefined structural components (Theme and Explanation parts) to enforce interpretability. Future work should explore more dynamic approaches that allow for adaptive sentence structuring based on different types of tables, potentially improving performance across varied tabular structures.

References

Soyoung An, Kyunghoon Bae, Eunbi Choi, Stanley Jungkyu Choi, Yemuk Choi, Seokhee Hong, Yeonjung Hong, Junwon Hwang, Hyojin Jeon, Gerrard Jeongwon Jo, Hyunjik Jo, Jiyeon Jung, Yountae Jung, Euisoon Kim, Hyosang Kim, Joonkee Kim, Seonghwan Kim, Soyeon Kim, Sunkyoung Kim, and 18 others. 2024. EXAONE 3.0 7.8b instruction tuned language model. *CoRR*, abs/2408.03541.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Changwook Jun, Jooyoung Choi, Myoseop Sim, Hyun Kim, Hansol Jang, and Kyungkoo Min. 2022. Korean-specific dataset for table question answering. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, Marseille, France, 20-25 June 2022*, pages 6114–6120. European Language Resources Association.

Tianyang Liu, Fei Wang, and Muhao Chen. 2024. Rethinking tabular data understanding with large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, pages 450–482. Association for Computational Linguistics.

Emanuele Musumeci, Michele Brienza, Vincenzo Suriani, Daniele Nardi, and Domenico Daniele Bloisi. 2024. LLM based multi-agent generation of semi-structured documents from semantic templates in the public administration domain. In *Artificial Intelligence in HCI*, pages 98–117, Cham. Springer Nature Switzerland.

Linyong Nan, Chiachun Hsieh, Ziming Mao, Xi Victoria Lin, Neha Verma, Rui Zhang, Wojciech Kryscinski, Hailey Schoelkopf, Riley Kong, Xiangru Tang, Mutethia Mutuma, Ben Rosand, Isabel Trindade, Renusree Bandaru, Jacob Cunningham, Caiming Xiong, and Dragomir R. Radev. 2022. Fetaqa: Free-form table question answering. *Trans. Assoc. Comput. Linguistics*, 10:35–49.

National Institute of Korean Language, NIKL. 2024. A corpus for evaluating the generation of interpretations for part of a table. https://kli.korean.go.kr/corpus/taskOrdtm/taskDownload.do?taskOrdtmId=109&clCd=ING_TASK&subMenuId.

OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.

Elias Stengel-Eskin, Kenton W. Murray, Sheng Zhang, Aaron Steven White, and Benjamin Van Durme. 2021. Joint universal syntactic and semantic parsing. *Trans. Assoc. Comput. Linguistics*, 9:756–773.

Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Miculicich, Yasuhisa Fujii, Jingbo Shang, Chen-Yu Lee, and Tomas Pfister. 2024. Chain-of-table: Evolving tables in the reasoning chain for table understanding. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.

Biao Zhang, Zhongtao Liu, Colin Cherry, and Orhan Firat. 2024a. When scaling meets LLM finetuning: The effect of data, model and finetuning method. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Tianshu Zhang, Xiang Yue, Yifei Li, and Huan Sun. 2024b. Tablellama: Towards open large generalist models for tables. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, pages 6024–6044. Association for Computational Linguistics.

Jun Zhao, Jingqi Tong, Yurong Mou, Ming Zhang, Qi Zhang, and Xuanjing Huang. 2024. Exploring the compositional deficiency of large language models in mathematical reasoning through trap problems. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 16361–16376. Association for Computational Linguistics.

Table title : 보건복지부 관서 운영비 한도 Ministry of Health and Welfare's office operating expense limits

예산과목 Budget Category	용 도 Usage
운영비(210) Operating Expenses (210)	건당 500만원이하의 경비(단, 공관금 및 위원회 참석비, 선거 관련 용품제작·인쇄 비용·우편요금, 청사임차료, 기업특별회계 당해 사업에 직접 소요되는 경비는 금액 제한 없음). Expenses of up to 5 million KRW per case (However, there is no spending limit for embassy fees, committee attendance fees, election-related item production/printing costs, postage fees, building rental fees, and
여비(220) Travel Expenses (220)	전체 Entirely covered
업무추진비(240) Business Promotion Expenses (240)	건당 500만원 이하의 경비(단, 기업특별회계의 당해 사업에 직접 소요되는 경비는 금액제한 없음) Expenses of up to 5 million KRW per case (However, there is no spending limit for expenses directly related to projects under the corporate special accounting).

highlighted_cells : [(0,1), (1,1)]

Generic prompt

위의 표에서 highlighted_cells를 모두 언급하여 내용을 요약해줘
Summarize the contents of the table above by mentioning all highlighted_cells in the table above.

운영비(210)는 500만 원 이하의 경비가 기업 특별회계 사업에 직접 소요되는 경비로 사용되며, 후자의 경우 금액 제한이 없습니다.

Operating expenses (210) are expenses of 5 million won or less and are used directly for corporate special accounting projects, and in the case of the latter, there is no limit to the amount.

Generic Prompt with Specific Instructions

위 표에서 highlighted_cells를 모두 언급하여 정확하고 객관적 사실만 담아서 내용을 요약해줘.
Summarize the content with only accurate and objective facts, while mentioning all highlighted_cells in the table

보건복지부 관서 운영비 한도에 따르면, 운영비(210)는 건당 5백만 원 이하로 제한되나 기획특별회계 사업 직접 경비는 제한이 없고, 여비(220)는 전액 적용됩니다.

According to the Ministry of Health and Welfare's office operating expense limits, operating expenses (210) are limited to 5 million won or less per case, but there is no limit to direct expenses for planning and special accounting projects, and travel expenses (220) are fully applied.

Journalist persona-applied prompt

너는 한 문장짜리 기사를 쓰는 신문 기자야. 이 표에서 highlighted_cells를 모두 언급하여 한 문장으로 기사를 써줘.
You are a newspaper journalist writing a one-sentence news article. Write a single sentence mentioning all the highlighted_cells in the table.

보건복지부 관서 운영비 한도에 따르면 운영비(210)는 공관금 및 위원회 참석비, 선거 관련 용품 제작·인쇄 비용·우편요금, 청사 임차료 등을 포함하여 건당 500만원 이하로 제한되며, 여비(220)는 전체적으로 적용된다.

According to the Ministry of Health and Welfare's office operating expense limits, operating expenses (210) are restricted to 5 million KRW per case, including embassy fees, committee attendance fees, election-related item production and printing costs, postage fees, and building rental fees, while travel expenses (220) are fully covered.

Figure 2: An example of how a persona-applied prompt generates more accurate summaries. (Gray text indicates the English translation.)

A Korean table interpretation benchmark


```
{
 "id": "nikluge-gtps-2023-train",
 "input": {
  "metadata": {
   "title": "Tax Environment Changes and Policy Issues due to the 4th Industrial Revolution",
   "table_title": "Major Legislative Proposals in Taxation Related to the 4th Industrial Revolution",
   "date": "2020-06-09",
   "publisher": "National Assembly Budget Office",
   "url": "https://www.nabo.go.kr/Sub/01/Report/01_01_Board.jsp",
   "highlighted_cells": [[10, 13], [1, 14], [3, 14]]
  },
  "table": [
   { "value": "Special Tax Limitation Act", "is_header": true, "col": 0, "colspan": 4, "row": 0, "rowspan": 1 },
   { "value": "2009580", "is_header": false, "col": 0, "colspan": 1, "row": 1, "rowspan": 1 },
   { "value": "Special Deduction/Registration Fee", "is_header": false, "col": 1, "colspan": 1, "row": 1, "rowspan": 1 },
   { "value": "Income Tax Act Disclosure", "is_header": false, "col": 1, "colspan": 1, "row": 1, "rowspan": 1 }
  ]
 },
 "output": [
  "The content of the VAT law issued on November 6, 2018, covers the electronic application scope for VAT...",
  "For VAT law, the scope of electronic application includes internet ads, cloud computing services, ...",
  "On November 6, 2018, the VAT law was revised to include internet ads, cloud computing services,..."
 ]
}
```


Figure 3: An example from the corpus for evaluating interpretation generation of table segments (originally in Korean, translated into English) (NIKL, 2024).

We leverage the Korean table interpretation benchmark provided by the National Institute of Korean Language (NIKL, 2024).

As shown in Figure 3, an objective of the dataset is to summarize the highlighted cells, which are labeled in the metadata as `highlighted_cells`, into a single coherent sentence.

B Examples of the Theme Part & Explanation Part

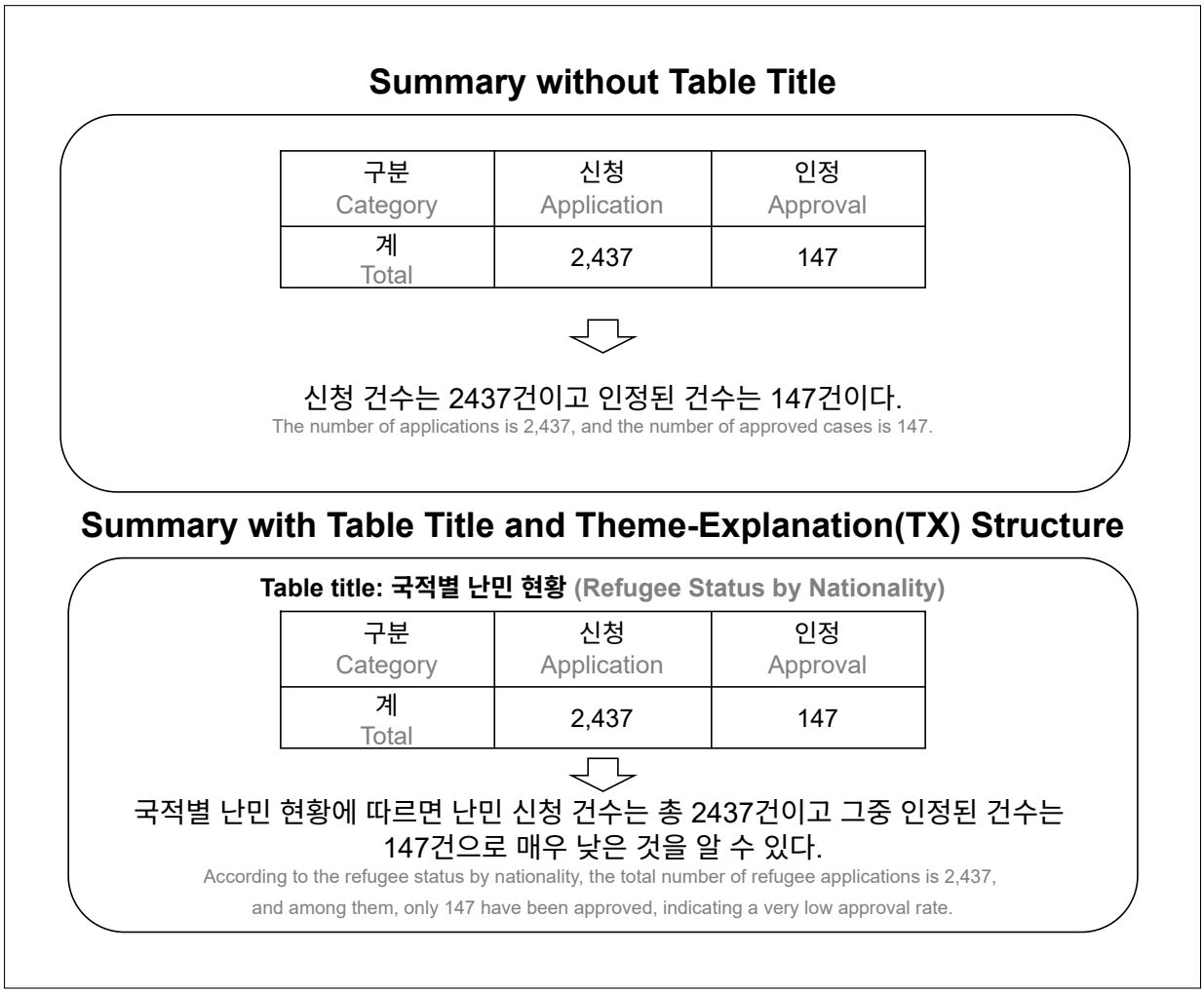

Figure 4: A sentence including the table title conveys the context more accurately. (Gray text indicates the English translation.)

B.1 Theme Part

For example, in the sentence: *"According to the refugee status by nationality, the total number of refugee applications is 2,437, and among them, only 147 have been approved, indicating a very low approval rate."* Here, the Theme Part is: *"According to the refugee status by nationality,"* This phrase, introduced with the citation expression *"According to"*, provides essential context for the numerical values that follow. Without this structured introduction, the reader may struggle to understand the significance of the numbers. Figure 4 illustrates how omitting the Theme Part results in an unclear or misleading summary.

B.2 Explanation Part

For instance, in the previous example, the Explanation Part is: *"the net fiscal cost increased by 9.435 trillion KRW from the previous year, reaching a total of 61.301 trillion KRW."* Here, the Explanation Part is derived by comparing the numerical changes between the two cells. Here, a trend analysis is applied to highlight the increase in fiscal cost.

C Implementation Details

C.1 Data Preprocessing

The first step is preprocessing the table to simplify its structure for better LLM comprehension. Since LLMs primarily operate on sequential text representations, directly processing raw tabular formats can lead to misinterpretation of hierarchical relationships within the data. To address this, we convert table data into a key-value pair dictionary format, which is commonly used in natural language processing tasks. This transformation significantly enhances LLMs' ability to recognize table semantics, improving summarization accuracy (Stengel-Eskin et al., 2021).

Then, we process merged cells to clarify the table structure. Merged cells span multiple rows or columns and are defined by 'rowspan' and 'colspan.' As shown in Figure 5, LLMs infer relationships between data through row or column alignment. However, incorrect handling of merged cell ranges can lead to misinterpretation. For example, in Figure 6, the cell labeled "2020" should cover columns 3 and 4, but it appears only in column 3. To resolve this, merged cells are replicated across their ranges, allowing LLMs to recognize cell dependencies and hierarchical structures correctly.

Finally, the transformed dictionary list retains only the highlighted and related cells, where "related cells" refer to all header cells sharing the same row/column as the highlighted cells. This process reduces data complexity and enhances LLMs' recognition of table structures.

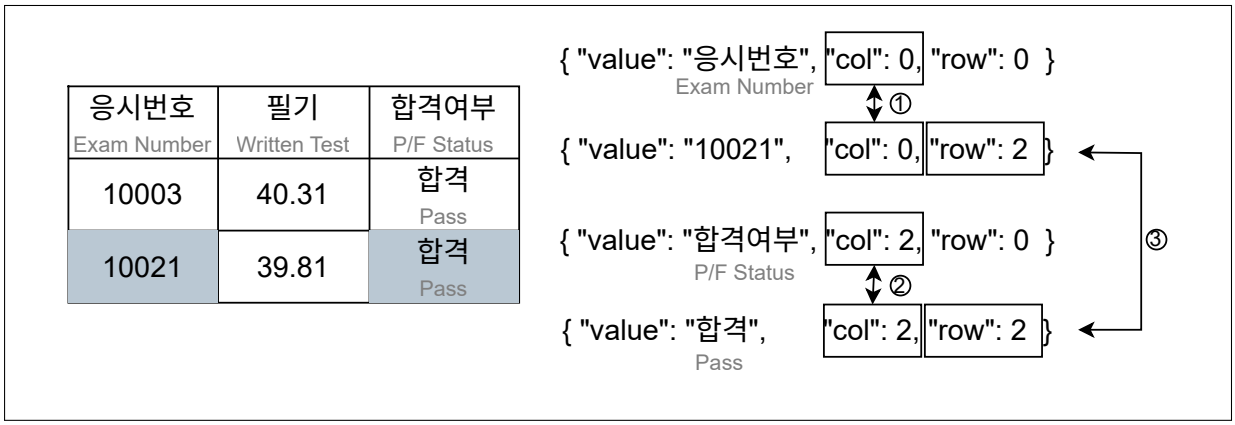

Figure 5: An example of inferring relationships between data sharing the same row or column. Through inference in ①, it is deduced that '10021' represents the 'Exam number.' In ②, the meaning of 'pass' is inferred. In ③, it is deduced that the exam with the 'Exam number' '10021' has 'passed.' (Gray text indicates the English translation.)

사업명 Project Name	2019 결산 2019 Settlement	2020	
		본예산 Main Budget	추경(A) Supplementary Budget(A)
고충민원조사활동 High-Rise Civil Complaint	1,286	1,295	1,295

{ 'row': 1, 'col': 1, 'rowspan': 2, 'colspan': 1, 'value': '사업명'}
{ 'row': 1, 'col': 2, 'rowspan': 1, 'colspan': 2, 'value': '2019 결산'}
{ 'row': 1, 'col': 4, 'rowspan': 1, 'colspan': 2, 'value': '2020'}
{ 'row': 2, 'col': 2, 'rowspan': 1, 'colspan': 1, 'value': '본예산'}
{ 'row': 2, 'col': 3, 'rowspan': 1, 'colspan': 1, 'value': '추경(A)'}
{ 'row': 3, 'col': 1, 'rowspan': 1, 'colspan': 1, 'value': '고층민원조사활동'}
{ 'row': 3, 'col': 2, 'rowspan': 1, 'colspan': 1, 'value': '1,286'}
{ 'row': 3, 'col': 3, 'rowspan': 1, 'colspan': 1, 'value': '1,295'}
{ 'row': 3, 'col': 4, 'rowspan': 1, 'colspan': 1, 'value': '1,295'}

고층민원조사활동 추경(A)가 1,295,000,000 원인 해의 연도를 알려줘

Tell me the year when the supplementary budget (A) for the High-Rise Civil Complaint Investigation Activities was 1,295,000,000 KRW.

이 표는 2019년도 2020년도 고층민원조사활동을 다루고 있지만 1,295,000,000 원인 해의 연도는 알 수 없습니다.

This table covers the High-Rise Civil Complaint Investigation Activities for 2019 and 2020, but it does not indicate the year in which the supplementary budget of 1,295,000,000 KRW was allocated.

Figure 6: An example of how merged cells hinder table recognition. (Gray text indicates the English translation.)

C.2 Prompt Details

Tabular-TX generates table summaries in two steps. First, the **Data Recognition / Classification** step identifies key data from highlighted cells (Figure 7). Second, the **Sentence Generation** step forms a summary in the Theme-Explanation format (Figure 8).

Title: {'title'}

Table Title: {'table_title'}

Table: {‘table data}

highlighted_cells: {'highlighted_cell coordinates'}

highlighted_cells are in (col, row) format and indicate the location of important data within the table.

if ’△’ in Table — [{', '.join(increased_cells)}] values contain the △ symbol, which indicates an increase.

if ’monetary unit’ in Table — [{', '.join(money_cells)}] values represent monetary amounts.

if ’percentage data’ in Table — [{', '.join(percent_cells)}] values represent percentages and should be displayed with a % symbol.

Figure 7: Summarizing key data points from the table in a single sentence for a news article. (originally in Korean, translated into English)

You are a newspaper reporter writing an article based on the table. You must convey the information in a single sentence. Mention all the highlighted_cells in the table and write the sentence in a declarative form. Do not say anything other than the one sentence.

Figure 8: Writing a one-sentence summary of a table by embodying news reporter persona. (originally in Korean, translated into English)