Understanding Tables in Financial Documents

Shared Tasks for Table Retrieval and Table QA on Japanese Annual Securities Reports

Yasutomo Kimura[1,*], Eisaku Sato[1], Kazuma Kadowaki[2] and Hokuto Ototake[3]

[1]*Otaru University of Commerce, Hokkaido, Japan*
[2]*The Japan Research Institute, Limited, Tokyo, Japan*
[3]*Fukuoka University, Fukuoka, Japan*

Abstract

This paper presents a framework for the "NTCIR-18 U4" and "SIG-FIN UFO-2024" shared tasks, which focus on tables within annual securities reports. Annual securities reports are critical documents that provide insights into a company's financial status and business performance. However, challenges remain in accurately and efficiently analyzing the data they contain. To address these issues, we propose two sub-tasks for the above shared tasks: Table Retrieval and Table QA tasks, which utilize datasets from TOPIX100 and TOPIX500 annual securities reports. Participants are tasked with developing systems (programs) that automatically process data for the two tasks and compete for top performance on a leaderboard. Accuracy scores and rankings are determined by submitting the task's output, in JSON format, to the leaderboard. Through these shared tasks, we aim to enhance the utility of annual securities reports and advance natural language processing technologies for financial data analysis.

Keywords

annual securities report, shared task, table retrieval, table question-answering

1. Introduction

In recent years, financial disclosures have become essential for investors seeking to make informed decisions based on reliable corporate data. In Japan, listed companies are required to submit an annual securities report, a statutory disclosure document that provides comprehensive information on business operations, financial data, risk factors, corporate governance, and shareholder information. These reports, accessible via the Electronic Disclosure for Investors' NETwork (EDINET)[1], serve as a critical information source for investors aiming to compare companies effectively.

These securities reports are structured in XBRL (eXtensible Business Reporting Language), an XML-based format designed to standardize and facilitate the production, distribution, and reuse of financial information. By incorporating "taxonomies" that define the structure and meaning of data, XBRL enables automated processing, potentially streamlining financial analysis.

However, practical challenges arise due to the presence of untagged data and the existence of unique taxonomies created by different report submitters, complicating the identification of comparable elements across reports.

To this end, we propose two tasks that aim to facilitate cross-company comparisons by focusing on the tables and text within annual securities reports. The first task is the NTCIR-18 U4 task, adopted by Japan's National Institute of Informatics (NII) as part of NTCIR-18[2]. The second is the SIG-FIN UFO-2024 task, organized by the Financial Informatics Study Group (SIG-FIN) under the Japanese Society for Artificial Intelligence (JSAI). The former focuses on TOPIX100 annual securities reports submitted between April 1, 2020, and March 31, 2021, while the latter focuses on TOPIX500 annual securities reports submitted between July 1, 2023, and June 30, 2024.

Question : For Bandai Namco Holdings Inc., what were the ``net assets and key management indicators'' as of 2020?
質問文：株式会社バンダイナムコホールディングスの2020年度時点における「純資産額、経営指標等」は？

TR; Table Retrieval

Input
1. Question
2. HTML of the Annual Securities Report

Bandai Namco
- 0000000.html
- 0101010.html
- 0102010.html

Output A "table" that contains the answers to the questions

Table QA; Table Question Answering

Input
1. Question
2. The "table" that contains the answers to the questions

| Fiscal Year | | 11th Term | 12th Term | 13th Term | 14th Term | 15th Term |
Fiscal Year End		March 2016	March 2017	March 2018	March 2019	March 2020
Net Sales	(Million Yen)	575,504	620,061	678,312	732,347	723,989
Ordinary Income	(Million Yen)	50,774	63,290	75,380	86,863	79,797
Net Income Attributable to Owners of Parent	(Million Yen)	34,583	44,159	54,109	63,383	57,665
Comprehensive Income	(Million Yen)	27,377	43,104	56,024	69,952	53,312
Net Assets	(Million Yen)	317,304	348,784	387,354	429,644	454,684
Total Assets	(Million Yen)	448,336	488,032	540,490	612,955	617,819

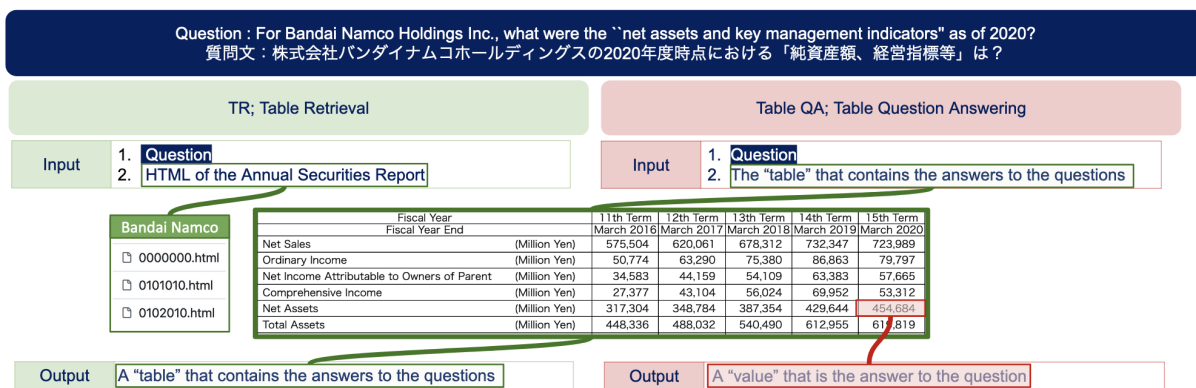**Output** A "value" that is the answer to the question

Figure 1: Overview of the Table Retrieval Task and Table QA Task

We organized these shared tasks in collaboration with the NII Testbeds and Community for Information Access Research (NTCIR) [1], which specializes in information retrieval, and the SIG-FIN group of the JSAI, which focuses on financial technology. Through these initiatives, we aim to attract researchers and practitioners interested in these fields and contribute to further advancing technologies at the intersection of finance and information retrieval.

In each shared task, we conducted two sub-tasks: Table Retrieval, which involves searching for tables, and Table Question Answering (Table QA), which involves answering questions by identifying the target cells within the tables, as illustrated in Figure 1. We designed each sub-task and constructed datasets for each task.

The contributions of this study are as follows:

- Design of two tasks, Table Retrieval and Table QA, targeting securities reports.
- Construction of datasets for Table Retrieval and Table QA, and their release on GitHub[3].
- Organization of the NTCIR-18 U4 task and the SIG-FIN UFO-2024 task.

2. Related Work

2.1. Research on Tables

A **table** is a data format with a two-dimensional structure used to organize and manage knowledge or information, and it is widely utilized in various contexts. However, not all tables have a highly structured database format; they are often represented as semi-structured data. Furthermore, the data contained in table cells is not limited to numerical values; it frequently includes strings and other non-numerical data. Numerous methods have been proposed to accommodate such diverse table data. Zhang and Balog [2] surveyed on tables on the web and classified approaches to accessing table data into six main categories.

1. Table extraction
2. Table interpretation
3. Table search
4. Table question answering
5. Knowledge base augmentation
6. Table augmentation

In addition, table-related tasks include table fact verification [3, 4], table detection (searching for tables within documents) [5, 6], spreadsheet manipulation [7], column type annotation [8], and entity linking (linking to knowledge bases) [9, 10]. These tasks are critical in information retrieval and data analysis based on table data, and they are particularly anticipated in fields where handling large-scale data and automation are required. Recently, approaches utilizing large language models (LLMs) and visual language models (VLMs) have been increasing, and research on learning methods, prompt engineering, and agents is also gaining attention [11]. Our shared tasks (NTCIR-18 U4 and the SIG-FIN UFO-2024) are related to table search, table detection, and table question answering (Table QA).

2.2. Table Retrieval and Table QA

Table retrieval aims to identify appropriate tables from vast datasets [12]. In this task, methods that typically assign relevance scores on the basis of the relationship between natural language queries and individual tables are commonly used.

Table QA refers to the technology that provides appropriate answers from tables in response to user questions. Approaches to Table QA include semantic parsing-based, generation-based, extraction-based, matching-based, and retrieval-based methods [13]. The difficulty in Table QA lies in the need to handle semi-structured or unstructured data, as it also involves non-database tables.

Compared to existing Tabular QA datasets such as FinQA[14] and TAT-QA[15], which primarily focus on English-language datasets and are designed to handle numerical reasoning in financial contexts, our proposed shared tasks specifically target the Japanese language. Japanese tabular and textual data often exhibit unique linguistic and structural features distinct from those in English datasets. These features may include variations in numerical data formats, context-dependent expressions, and implicit relational cues.

2.3. Tables in the Financial Domain

Hybrid data, which includes both tables and text, such as in financial reports, is quite prevalent in the real world [16]. Zhu et al. constructed a question-answering benchmark dataset focused on the hybrid content of tabular and textual data in the financial domain [15].

Pan et al. proposed CLTR, an architecture for end-to-end table retrieval at the cell level [17]. While CLTR can be applied to open-domain datasets, including finance and healthcare ones, its performance specifically within the financial domain has not been clarified, nor does it target the Japanese language.

One of the tasks focused on Japanese financial table structure analysis is the UFO (Understanding of non-Financial Objects in Financial Reports) task [18]. The UFO task aims to extract structured information from tables and text found in annual securities reports and consists of two sub-tasks: the Table Data Extraction (TDE) task and the Text-to-Table Relationship Extraction (TTRE) task. The TDE task classifies cells in tables into four categories with the goal of identifying the type of each cell: metadata, header, attribute, and data [19]. The main focus of TDE was on cell classification, and additional processing to enable inter-company comparisons remained unexplored.

3. NTCIR-18 U4 and SIG-FIN UFO-2024 Tasks

Both the NTCIR-18 U4 and SIG-FIN UFO-2024 tasks consist of two sub-tasks: the **Table Retrieval** task, which involves searching for tables, and the **Table Question Answering (Table QA)** task, which involves answering questions by identifying the target cells within the tables [20]. Figure 1 illustrates the concept of these two sub-tasks.

3.1. Table Retrieval: Table Search Task

Table Retrieval is a task that involves searching for a "table" containing the values that answer a given question from the tables included in a company's annual securities report. On average, a company's

annual securities report contains 221.9 tables [21], and it is necessary to identify the specific table that contains the answer to the question needs to be identified. The input, output, and evaluation criteria for this task are as follows.

Input	1. Question
	2. HTML file of the annual securities report
Output	Table (Table ID)
Evaluation	Accuracy

An example of input and output is shown below.

Input	1. For Bandai Namco Holdings Inc.,
	what were the "net assets and key management indicators" as of 2020?
	2. S100ISF1-0000000.html, S100ISF1-0101010.html, ...
Output	S100ISF1-0101010-tab2

For the input, HTML files downloaded from EDINET are used. Each table element (`<table>`) in the HTML files is assigned a unique Table ID, and when outputting the table that answers the question, this Table ID is used as the output. In the output example above, the Table ID "S100ISF1-0101010-tab2" refers to the second table in the "S100ISF1-0101010.html" file, with "-tab{*table number*}" appended to the file name.

The metric used for evaluation is accuracy, which is calculated by dividing the number of correct outputs by the total number of inputs in the test dataset.

We evaluated a few baseline methods using our validation datasets, which contain 3,131 and 1,533 questions for NTCIR-18 U4 and SIG-FIN UFO-2024, respectively. The results are shown in Table 1. For the NTCIR-18 U4 task, the highest accuracy of 0.2111 was achieved by using the text-embedding-3-small model to create embeddings based on Cell Text. Similarly, for the SIG-FIN UFO-2024 task, a top accuracy of 0.1937 was obtained using the text-embedding-3-large model for Cell Text embeddings.

Table 1
Validation Results for the Table Retrieval Task

Task	Baseline methods	Accuracy
NTCIR-18 UFO	text-embedding-3-small + Cell Text	**0.2111**
	text-embedding-3-large + Cell Text	0.1833
	text-embedding-3-small + HTML Text	0.1843
	text-embedding-3-large + HTML Text	0.1418
	text-embedding-3-small + Markdown Text	0.1233
	text-embedding-3-large + Markdown Text	0.1383
SIG-FIN UFO-2024	text-embedding-3-small + Cell Text	0.1657
	text-embedding-3-large + Cell Text	**0.1937**

3.2. Table QA: Table Question Answering Task

The Table QA task, given a target table, identifies the "value" that answers the question. To accurately determine the answer, complex tables included in annual securities reports need to be handled [22]. The input, output, and evaluation criteria for this task are as follows:

Input	1. Question
	2. Target table (Table ID)
	3. HTML file of the annual securities report
Output	Value, Cell ID
Evaluation	Accuracy (value), Accuracy (cell ID)

An example of input and output is shown below.

Input	1. For Bandai Namco Holdings Inc.,
	what were the "net assets and key management indicators" as of 2020?
	2. S100ISF1-0101010-tab2
	3. S100ISF1-0000000.html, S100ISF1-0101010.html, ...
Output	S100ISF1-0101010-tab2-r8c1

For input data, HTML files and Table IDs are provided, allowing the system to extract the range enclosed by the `<table>` tag from the HTML file. Additionally, if necessary, the system can utilize the surrounding context of the table.

Similar to Table IDs, each cell (`<th>` and `<td>` tags) within the table in the HTML file is assigned a unique Cell ID. When outputting the value corresponding to the answer to the given question, this Cell ID is used. In the output example above, the Cell ID is the Table ID of the table containing the cell, with "-r{*row number*}c{*column number*}" appended, so the Cell ID "S100ISF1-0101010-tab2-r8c1" refers to the cell in the 8th row and 1st column of the table "100ISF1-0101010-tab2".

For evaluation, similar to the Table Retrieval task, accuracy is calculated by dividing the number of correct outputs by the total number of inputs in the test dataset. However, discrepancies between the value contained in the HTML cell and the expected answer are frequently observed. For example, if the expected answer is "4448000000", the corresponding cell in the HTML might contain the string "4,448", while another cell, such as in the top right or column name, might indicate "(in millions of yen)". In this case, the system answering the task must reference both cells to generate the answer "4,448 million yen". While this is equivalent to the correct answer, to compare it accurately, the system must replace the string "million yen" with "000000" and remove the comma.

Due to this, in this task, both the response and the correct answer are normalized before calculating accuracy. The normalization specification was continually revised during the "dry run" period, considering feedback from participants.

We evaluated a few baseline methods using our validation datasets, which contain 3,132 and 1,534 questions for NTCIR-18 U4 and SIG-FIN UFO-2024, respectively. These baseline methods involved converting the target table into text format and inputting it, along with the question, into an LLM to generate the desired values. The results are shown in Table 2. For the NTCIR-18 U4 task, the highest accuracy of 0.7471 was achieved by using the Claude 3 Opus model. Similarly, for the SIG-FIN UFO-2024 task, a top accuracy of 0.5750 was obtained using the GPT-4o model.

4. Dataset

4.1. Securities Reports Used in Our Dataset

The NTCIR-18 U4 task focuses on analyzing securities reports from companies in the TOPIX100 index. The dataset consists of securities reports from companies that are part of the TOPIX100, submitted between April 1, 2020, and March 31, 2021.

The SIG-FIN UFO-2024 task, on the other hand, focuses on analyzing securities reports from companies in the TOPIX500 index. The annual securities reports used in this task are drawn from the TOPIX500, which represent publicly listed companies with high market capitalization and liquidity. For this task, we target the annual securities reports of 497 companies[4] constituting the TOPIX 500 as of April 30,

---

[4]Despite the name TOPIX500, as of 30 April 2024, it only includes 497 companies. To be more precise, the TOPIX500 is a

Table 2
Validation results for the Table QA task

Task	Baseline methods	Accuracy (value)
NTCIR-18 U4	GPT-4o (gpt-4o-2024-05-13)	0.6475
	GPT-3.5-turbo (gpt-3.5-turbo-0125)	0.3493
	Gemini 1.5 Pro (gemini-1.5-pro-001)	0.5744
	Gemini 1.5 Flash (gemini-1.5-flash-001)	0.4898
	Claude 3 Opus (claude-3-opus-20240229)	**0.7471**
	Claude 3 Haiku (claude-3-haiku-20240307)	0.3209
	Claude 3.5 Sonnet (claude-3-5-sonnet-20240620)	0.7216
SIG-FIN UFO-2024	GPT-4o (gpt-4o-2024-05-13)	**0.5750**
	GPT-3o-mini (gpt-4o-mini-2024-07-18)	0.3957

2024. The dataset includes 494 financial statements submitted to EDINET between July 1, 2023, and June 30, 2024.

To account for differences in the structure of annual securities reports across industries, we ensure that the dataset is balanced by industry. The annual securities reports are distributed across the training, validation, and test sets with minimal industry bias. Specifically, we use the ten major categories from the Tokyo Stock Exchange's 33 industry classifications (service industry, transportation and communications, finance and insurance, construction, mining, commerce, fisheries, agriculture and forestry, manufacturing, electricity and gas, and real estate). The data is divided such that the ratio of train:validation:test is approximately 6:1:3 within each industry category. This results in 289 companies' reports being used for training, 52 for validation, and 153 for testing.

We retrieve the financial data using the EDINET API v2, utilizing the XBRL, HTML, and CSV files available through the API. The XBRL files contain tabular data, such as taxonomies and instances, referred to as "XBRL information" below, which is also embedded in the corresponding HTML files. The CSV files, referred to below as "annual securities report CSVs," provide a more accessible format for the XBRL data for easier handling in the study.

4.2. Question Creation

Questions are created using annual securities report CSVs and question templates. In the annual securities report CSV, each row represents data, and each column shows the corresponding XBRL information (element ID, item name, context ID, relevant year, consolidated or individual, period or point in time, unit ID, unit, value). Among this XBRL information, the element ID and context ID are crucial for data extraction. The element ID indicates what the data represents, but it is not unique within a single annual securities report. Therefore, combining the element ID with the context ID, which represents the period and dimension, enables data within a report to be uniquely identified and the desired information to be extracted. Thus, the question must include both element ID and context ID.

On the basis of this, the initial version of the question is created as follows:

> ┌─ Question (Initial Version) ──────────────────────────────
>
> What is the value of "{*Element ID*}" for {*Company Name*} in {*Context ID*}?

However, if the element ID and context ID are used as they appear in the annual securities report CSV, the question will not be meaningful in Japanese as they are simply IDs. Therefore, the context ID is represented using the relative year, consolidated or individual, and period or point in time, while the element ID is expressed as the item name. The final version of the question is defined as follows:

---

stock price index composed of the TOPIX Core30, TOPIX Large70 and TOPIX Mid400, but of these, only 397 companies are included in the TOPIX Mid400.

> **Question (Detailed Version)**
>
> What is the value of "{*Item Name*}" in the {*Year*} {*Period or Point in Time*} {*Consolidated or Individual (optional)*} annual securities report of {*Company Name*} for {*Member Element (optional)*}?

The explanations for each part are as follows:

- Year: Calculated based on the basis of the relevant year, and the string is included in the question.
- Period or Point in Time: If it is a point in time, the word "point" is added right after the year.
- Consolidated or Individual: If it is consolidated or individual, the corresponding string is included; otherwise, "annual securities report of" is omitted.
- Member Element: If the context ID contains a member element, the string is included. This element is not translated into Japanese to ensure uniqueness and is used as-is from the annual securities report CSV (ensuring uniqueness is a future challenge).
- Item Name: This is essentially the Japanese translation of the element ID, so the string is included.

An example of a question created using the template is as follows:

> **Example Created with Question Template**
>
> What is the value of "Building (net amount)" in the 2020 individual annual securities report of Daiwa House Industry Co., Ltd. for `NonConsolidatedMember`?

When creating questions for the SIG-FIN UFO-2024 dataset, we also performed data sampling to avoid generating too many similar questions. For data sampling, a unique list of item names is created for each company, and random sampling is performed so that 1/10th of the entire dataset is selected. Additionally, the number of samples per item name is adjusted on the basis of the number of data entries for each item name[5].

As a result of these procedures, we constructed the NTCIR-18 U4 dataset consisting of 32,587 entries for the Table Retrieval task and 32,589 entries for the Table QA task, and the SIG-FIN UFO-2024 dataset consisting of 14,410 entries for the Table Retrieval task and 14,412 entries for the Table QA task, as shown in Table 3[6].

Table 3
Breakdown of Each Dataset for Dry Run

Task		Train	Validation	Test	Total
NTCIR-18 U4	Table Retrieval	22,982	3,131	6,474	32,587
	Table QA	22,982	3,132	6,475	32,589
SIG-FIN UFO-2024	Table Retrieval	8,390	1,533	4,487	14,410
	Table QA	8,390	1,534	4,488	14,412

5. Schedule

To encourage broad participation from those interested in finance, the organizers have introduced two complementary tasks: the NTCIR-18 U4 and the SIG-FIN UFO-2024. The SIG-FIN community includes researchers and practitioners actively engaged in finance, while NTCIR attracts participants interested in shared tasks, especially those with a focus on information retrieval and natural language processing.

---

[5]If there is only one data entry for an item name, one sample is taken; if there are two to five, two samples are taken; if there are six or more, three samples are randomly selected. Data with submitter-specific taxonomies that do not include an item name in the annual securities report CSV, or data not from tables (i.e., data not in HTML's `<td>` tags) are excluded.

[6]These are the breakdowns of the initial datasets used for the Dry Run. The datasets used in the Formal Run phase have been modified to fix a couple of issues, and as a result, they contain a slightly different number of entries.

By running similar tasks across these two distinct communities, we aim to foster interaction among participants with diverse expertise and perspectives on finance, creating an opportunity for knowledge exchange and collaboration. We look forward to welcoming a diverse group of participants to build a comprehensive and impactful competition.

The schedule for each task is outlined in Table 4, showing the parallel timelines and key phases for both NTCIR-18 U4 and SIG-FIN UFO-2024. As illustrated in the table, both tasks share similar phases such as a dry run, formal run, and evaluation period, which will allow participants to apply and refine their approaches across both tasks seamlessly. This alignment ensures that participants will have the opportunity to benefit from complementary insights across the two tasks and fosters collaborative learning within the community.

Table 4
Timeline for NTCIR-18 U4 and SIG-FIN UFO-2024 Tasks

Task	Phase	NTCIR-18 U4 Task	SIG-FIN UFO-2024 Task
Preparation	Dataset Release	July 20, 2024	August 15, 2024
Initial Briefing	Online Session	July 20, 2024	-
Dry Run		July 20–October 31, 2024	August 15 – October 31, 2024
Formal Run		November 1–December 28, 2024	November 1–December 28, 2024
Evaluation	Results Return	February 1, 2025	Mid-January, 2025
Publication	Paper Submission	May 1, 2025	Mid-February, 2025
Presentation	Final Conference	June 10–13, 2025	Early March, 2025

5.1. NTCIR-18 U4 Task

The schedule for the NTCIR-18 U4 task is as follows: The dataset for the NTCIR-18 U4 shared task was released in July 2024, followed by an online briefing session on July 20, 2024, where participants received essential information about the task. The dry run phase ran from July 2024 to October 31, 2024, during which participants worked on the dataset and refined their methods. Any issues identified in the dataset during this phase were addressed and resolved to ensure a smooth formal run. The formal run phase is scheduled from November 1, 2024, to December 28, 2024. Throughout the NTCIR-18 U4 task, a leaderboard will be used to provide participants with real-time feedback on their performance. Similar to the SIG-FIN UFO-2024 task, the NTCIR-18 U4 leaderboard will display a Public score on the basis of a subset of the test data during the task period, allowing participants to gauge their progress. Evaluation results and final rankings are scheduled to be returned to participants on February 1, 2025, along with a partial publication of the task overview paper summarizing key outcomes.

5.2. SIG-FIN UFO-2024 Task

The schedule for the SIG-FIN UFO-2024 task is as follows: The dataset for this shared task was released on August 15, 2024, with the dry run phase extending until October 31, 2024. During this phase, participants worked on developing their methods using the dataset, and any data issues identified during this period were addressed and corrected. The formal run phase is scheduled from November 1, 2024, to December 28, 2024.

The shared task ranking will be determined on the basis of the evaluation method used in Kaggle[7], incorporating both Public and Private scores. Throughout the shared task, the Public score (calculated from a subset of the test data) will be displayed on the leaderboard. After the shared task concludes, the Private score (evaluated on the remaining portion of the test data) will be calculated. The final results, based on the Private score, are scheduled to be announced at the 34th SIG-FIN in March 2025.

---

[7] https://www.kaggle.com/

6. Conclusion

This paper proposed a framework for two shared tasks, NTCIR-18 U4 and SIG-FIN UFO-2024, which focus on tables within annual securities reports. In these shared tasks, two sub-tasks are conducted: Table Retrieval and Table Question Answering (Table QA), which target the annual securities reports of companies belonging to the TOPIX 100 or TOPIX 500 indexes.

Acknowledgments

References

[1] T. Sakai, D. W. Oard, N. Kando, Evaluating Information Retrieval and Access Tasks: NTCIR's Legacy of Research Impact, The Information Retrieval Series, Springer Nature, 2021. doi:10.1007/978-981-15-5554-1.

[2] S. Zhang, K. Balog, Web table extraction, retrieval, and augmentation: A survey, in: ACM Transactions on Intelligent Systems and Technology (TIST), volume 11, issue 2, 2020, pp. 1–35. doi:10.1145/3372117.

[3] W. Chen, H. Wang, J. Chen, Y. Zhang, H. Wang, S. Li, X. Zhou, W. Y. Wang, TabFact: A large-scale dataset for table-based fact verification, in: 8th International Conference on Learning Representations, ICLR 2020, 2020. URL: https://openreview.net/forum?id=rkeJRhNYDH.

[4] F. Wang, K. Sun, J. Pujara, P. Szekely, M. Chen, Table-based fact verification with salience-aware learning, in: Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 4025–4036. doi:10.18653/v1/2021.findings-emnlp.338.

[5] L. Chen, C. Huang, X. Zheng, J. Lin, X. Huang, TableVLM: Multi-modal pre-training for table structure recognition, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 2437–2449. doi:10.18653/v1/2023.acl-long.137.

[6] D. Prasad, A. Gadpal, K. Kapadni, M. Visave, K. Sultanpure, CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 2439–2447. doi:10.1109/CVPRW50498.2020.00294.

[7] Z. Ma, B. Zhang, J. Zhang, J. Yu, X. Zhang, X. Zhang, S. Luo, X. Wang, J. Tang, SpreadsheetBench: Towards challenging real world spreadsheet manipulation, 2024. doi:10.48550/arXiv.2406.14991.

[8] P. Li, Y. He, D. Yashar, W. Cui, S. Ge, H. Zhang, D. Rifinski Fainman, D. Zhang, S. Chaudhuri, Table-GPT: Table fine-tuned GPT for diverse table tasks, in: Proceedings of the ACM on Management of Data, volume 2, issue 3, 2024, pp. 1–28. doi:10.1145/3654979.

[9] X. Deng, H. Sun, A. Lees, Y. Wu, C. Yu, TURL: table understanding through representation learning, in: Proceedings of the VLDB Endowment, volume 14, issue 3, 2020, p. 307–319. doi:10.14778/3430915.3430921.

[10] T. Zhang, X. Yue, Y. Li, H. Sun, 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), 2024, pp. 6024–6044. doi:10.18653/v1/2024.naacl-long.335.

[11] W. Lu, J. Zhang, J. Fan, Z. Fu, Y. Chen, X. Du, Large language model for table processing: A survey, 2024. doi:10.48550/arXiv.2402.05121.

[12] A. S. Sundar, L. Heck, cTBLS: Augmenting large language models with conversational tables, in:

Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023), 2023, pp. 59–70. doi:10.18653/v1/2023.nlp4convai-1.6.

[13] N. Jin, J. Siebert, D. Li, Q. Chen, A survey on table question answering: Recent advances, in: Knowledge Graph and Semantic Computing: Knowledge Graph Empowers the Digital Economy, Springer Nature Singapore, 2022, pp. 174–186. doi:10.1007/978-981-19-7596-7_14.

[14] Z. Chen, W. Chen, C. Smiley, S. Shah, I. Borova, D. Langdon, R. Moussa, M. Beane, T.-H. Huang, B. Routledge, W. Y. Wang, FinQA: A dataset of numerical reasoning over financial data, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 3697–3711. doi:10.18653/v1/2021.emnlp-main.300.

[15] F. Zhu, W. Lei, Y. Huang, C. Wang, S. Zhang, J. Lv, F. Feng, T.-S. Chua, TAT-QA: A question answering benchmark on a hybrid of tabular and textual content in finance, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 3277–3287. doi:10.18653/v1/2021.acl-long.254.

[16] N. Romanus Myrberg, S. Danielsson, Question-Answering in the Financial Domain, Master's thesis, Department of Computer Science, Lund University, 2023. URL: http://lup.lub.lu.se/student-papers/record/9126226.

[17] F. Pan, M. Canim, M. Glass, A. Gliozzo, P. Fox, CLTR: An end-to-end, transformer-based system for cell-level table retrieval and table question answering, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, 2021, pp. 202–209. doi:10.18653/v1/2021.acl-demo.24.

[18] Y. Kimura, T. Kondo, K. Kadowaki, M. Kato, UFO: Proposal for an information extraction task for tables in annual securities reports (in Japanese), in: JSAI Technical Report, Type 2 SIG, volume FIN-029, The Japanese Society for Artificial Intelligence, 2022, pp. 32–38. doi:10.11517/jsaisigtwo.2022.FIN-029_32.

[19] K. Kadowaki, Y. Kimura, M. Kato, T. Kondo, H. Ototake, Toward the construction of a dataset for table structure analysis for annual securities reports (in Japanese), in: JSAI Technical Report, Type 2 SIG, volume FIN-030, The Japanese Society for Artificial Intelligence, 2023, pp. 100–105. doi:10.11517/jsaisigtwo.2023.FIN-030_100.

[20] E. Sato, Y. Kimura, Creating a question-answering dataset for securities reports and evaluation of the method using LLM (in Japanese), in: IEICE Technical Report, volume 124, no. 173, The Institute of Electronics, Information and Communication Engineers, 2024, pp. 93–98. URL: https://www.ieice.org/publications/search/summary.php?id=132450&tbl=ken&lang=jp.

[21] E. Sato, Y. Kaji, Y. Kimura, Analysis of tabular data contained in the TOPIX100 annual securities report (in Japanese), The 21st Forum on Information Technology (FIT2022) E-021 (2022). URL: https://www.ieice.org/publications/conferences/summary.php?id=FIT0000015362&ConfCd=F&conf_type=F&year=2022.

[22] K. Okuyama, Y. Kimura, Analysis of machine-unreadable table structures in securities reports (in Japanese), The 30th Annual Meeting of the Association for Natural Language Processing (NLP2024) P3-20 (2024). URL: https://www.anlp.jp/proceedings/annual_meeting/2024/pdf_dir/P3-20.pdf.