NTCIR 17 Conference: Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, December 12-15, 2023, Tokyo, Japan
DOI: https://doi.org/10.20736/0002001321

Overview of NTCIR-17 UFO Task

Yasutomo Kimura
Otaru University of Commerce
RIKEN
Japan
kimura@res.otaru-uc.ac.jp

Hokuto Ototake
Fukuoka University
Japan
ototake@fukuoka-u.ac.jp

Kazuma Kadowaki
The Japan Research Institute, Limited
Japan
kadowaki.kazuma@jri.co.jp

Takahito Kondo
Kyoto Sangyo University
Japan
k4769@cc.kyoto-su.ac.jp

Makoto P. Kato
University of Tsukuba
Japan
mpkato@acm.org

ABSTRACT

The goal of the NTCIR-17 UFO task is to develop techniques for extracting structured information from tabular data and documents, focusing on annual securities reports. The Non-Financial Objects in Financial Reports (UFO) task consists of two subtasks: table data extraction (TDE) and text-to-table relationship extraction (TTRE). The TDE subtask, for understanding the structure of tables in annual securities reports, classifies each cell into one of four classes. The TTRE subtask is for linking the values of the tables with a relevant sentence in the text. We present the data used for and the results of the formal run for these subtasks.

TEAM NAME

Task Organizers

SUBTASKS

Overview

1 INTRODUCTION

Annual securities reports serve as an invaluable source of information that assists investors in identifying promising stocks from listed companies and in predicting the financial outcomes of their investments. An increasing number of investors have incorporated non-financial information, such as initiatives toward Sustainable Development Goals, corporate governance-related information, and business risks in the industry, as significant factors influencing investment decisions, and the volume of such information in securities reports is increasing every year.

Annual securities reports are written in XBRL, an XML-based language, so that financial information contained in income statements, balance sheets, and statements of cash flows can be easily extracted. However, many taxonomies in XBRL regarding the above-mentioned non-financial information are only defined in block units such as paragraphs. These blocks are described in free-format text with tables and figures in each company's format. In particular, the tables are not standardized in the scope of content each represents nor in the structure of the columns.

Therefore, to understand the structure of each table as a first step in dealing with such tables, we provide two subtasks, i.e., table data extraction (TDE) and text-to-table relationship extraction (TTRE), in the NTICR-17 Non-Financial Objects in Financial Reports (UFO) task.

2 RELATED WORK

There has been growing interest in applying natural-language-processing techniques to financial documents. FinNum-2 [2] is a task for fine-grained numeral understanding in financial social media data, and the Numeral Attachment subtask identifies the attached target of each numeral. FinCausal 2020 [3] is a shared task to identify causality in financial datasets. Bentabet et al. [1] organized a shared task at the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation (FNP-FNS 2020). The shared task, for extracting a table of contents (TOC) from investment documents, detected the document titles and organized them hierarchically into a TOC.

3 ANNUAL SECURITIES REPORTS

Annual securities reports contain information from financial statements, as well as an overview of the company and the state of its business. An annual securities report contains approximately 200 to 300 pages and consists of two parts: Part I of corporate information and Part II of information on the auditor company. The first part is typically organized as follows: Chapter 1: 'Corporate Overview', Chapter 2: 'Business Situation', Chapter 3: 'Facilities', Chapter 4: 'Status of the Submitting Company', Chapter 5: 'Financial Situation', Chapter 6: 'Outline of the Submitting Company's Share Administration' and Chapter 7: 'Reference Information of the Submitting Company'. Each chapter is written in XBRL, and an XML tag is assigned to each financial element in an income statement, balance sheet and cash flow statement. Of all the tables in Part I, financial information is tagged in about 18% of the tables, and even in Chapter I, 'Corporate Overview', which has the highest rate, only about 31% of the tables are assigned an XBRL tag. In this task, we carried out the annual securities reports from TOPIX 100 companies and conducted annotation by targeting the tables included in Chapters 1 to 6 of Part I.

4 TASK DEFINITION

We proposed the TDE and TTRE subtasks for this task.

4.1 TDE

4.1.1 *Purpose.*

Understanding the structure of tables in securities reports is a first step to enable a user (typically the investors) to extract the correct combination of items and values then compare the values

across different companies. Such a technology could also be used in applications such as question-answering, implication relation recognition, and fact-checking using tables.

The TDE subtask is performed to understand the structure of tables to deal with the tables contained in securities reports. Specifically, each table cell in a report is classified into four classes: Metadata, Header, Attribute, and Data. For the table shown in Figure 1, the company name (i.e., the primary key) is listed in each row, and attributes such as name and capital are listed in each column, and classification can be done after recognizing this structure. Each class is defined so that a statement of the form "Regarding the [Status of affiliated companies]$^{\text{Metadata}}$, the [capital]$^{\text{Header}}$ of [ABC, Ltd.]$^{\text{Attribute}}$ is [100 million yen]$^{\text{Data}}$" can be structured when the classification is conducted.

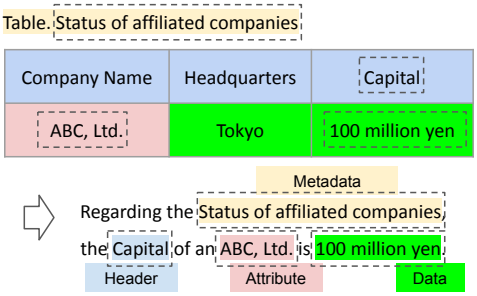

Figure 1: Annotation for TDE subtask

4.1.2 *Data.*

Data Source. We chose 20 companies listed in TOPIX 100, and three annotators annotated all the tables in their annual securities reports in HTML format. In this dataset, we used all the cells that were successfully annotated. That is, highly complex tables that cannot be described in the statement form "Regarding the [Metadata], the [Header] of an [Attribute] is [Data]." were excluded. As a result, 44.5% of all the tables included in the securities reports were used in the task.

Input. Annual securities reports in HTML format, where each cell to be classified is given IDs. Examples of the data input and output are shown in the Appendix C.1.

Output. A class of each cell, either Metadata, Header, Attribute or Data.

Data size. Table 1 shows the statistics for the TDE dataset for the formal run. Our dataset comprises 252 and 190 files and classifies 66,369 and 45,499 cells for the training and test, respectively.

4.1.3 *Evaluation.*

We used the accuracy of the quaternary classification task (i.e., Metadata, Header, Attribute, and Data) to evaluate the submitted systems.

4.1.4 *Baseline System.*

We developed a rule-based system that classifies the first row as Header, the first column (except the first row) as Attribute, and all remaining cells as Data. This system achieved an accuracy of 0.7981 in the formal run dataset and was provided as a baseline to the participants.

Table 1: TDE dataset statistics

Number of	Train	Test
Companies[a]	13	8
Securities reports	25	15
Chapters	252	190
Tables (in securities reports)	5,683	3,716
Cells (in securities reports)	247,302	173,732
Tables (used in the task)	2,530	1,660
Cells (used in the task)	66,369	45,499
Header	13,930	8,949
Attribute	11,236	7,611
Data	41,060	28,895
Metadata	143	44

[a]The total number of companies does not add up to 20 because securities reports from different years for one of the 20 companies were used in the Train and Test datasets. There is no overlap of security reports in the Train and Test, and hence of chapters, tables and cells.

4.2 TTRE

4.2.1 *Purpose.*

Linking text and tables plays an important step in finding the grounds for each claim and opinion in the securities reports, especially for numerical values that are often described in the form of tables. The TTRE subtask is performed to find table cells that are relevant to the main text of the securities reports. Each linked cell is also classified into an item name or a numerical value.

Figure 2 shows an example application of this subtask. When the orange text, "The dividend per share of 80.00 yen for the fiscal year ended March 31, 2016", is given (i.e., clicked), the corresponding table cells are highlighted. That is, the cells "The 77th (2016)" and "Dividend per share (yen)" are classified as Names (i.e., highlighted in red), while the cell "80.00" is classified as Value (i.e., highlighted in green).

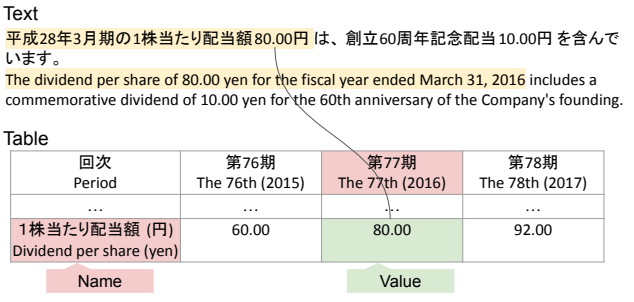

Figure 2: Example of TTRE subtask

4.2.2 *Data.*

We chose eight companies listed in TOPIX 100 and annotated all the 2,851 tables in their annual securities reports. Annotators first extracted text spans (usually noun phrases, e.g., 'Net sales') that

identify a specific cell in the same chapter and then linked with the cell. Each extracted phrase was given an ID. Each linked cell was then classified as **Value** if they were values such as monetary amount or percentage rates, **Name** if they described the **Value**, or **Etc** otherwise.

Input. Annual securities reports in HTML format, where each phrase to be linked and each table cell (i.e., link target candidate) are given IDs. Examples of the data input and output are shown in the Appendix C.2.

Output. Cell IDs relevant to each phrase and a class of either **Name** or **Value** for each selected cell.

Data size. Table 2 lists the statistics for the TTRE dataset for the formal run. The dataset is divided into two parts: Training and Test. For the Training set, there are 3,402 phrases, 1,726 tables, and 80,644 cells. The Test set consists of 1,875 phrases, 1,125 tables, and 47,517 cells. Thus, the TTRE dataset comprises 5,277 phrases, 2,851 tables, and 128,161 cells.

Table 2: TTRE dataset statistics

	Phrases	Tables	Cells
Training	3,402	1,726	80,644
Test	1,875	1,125	47,517
Total	5,277	2,851	128,161

4.2.3 *Evaluation.*

The TTRE subtask assigns several linking Name and Value cells to each phrase contained in the evaluation dataset. For each of Name and Value, precision, recall, and F1 values defined by the following equations were first calculated for each phrase:

$$\text{Precision} = \frac{\text{Number of cells with correct output}}{\text{Number of cells output}}$$

$$\text{Recall} = \frac{\text{Number of cells with correct output}}{\text{Number of cells in gold standard data}}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The macro average of the Name and Value scores shown in (1) and (2) below is then used as the score to evaluate the performance of each submitted system. In addition, the average of (1) and (2) was used as the overall score (3).

(1) Macro average of F1 score of Name of all phrases.
(2) Macro average of F1 scores of Value of all phrases.
(3) Average of (1) and (2).

4.2.4 *Baseline System.*

We did not provide any baseline system for the TTRE subtask. However, for reference to implement the input/output for the dataset, participants were provided with a sample program that outputs the randomly linked cell IDs.

5 SCHEDULE

The following is the schedule of the NTCIR-17 UFO task:

Preparation
November 8, 2022: UFO first round table meeting (online)
April 30, 2023: Initial release of the dataset
May 13, 2023: UFO second round table meeting (hybrid)
June 17, 2023 UFO third round table meeting (online)

Dry run
April 30—July 2, 2023: Dry run

Formal run
July 3, 2023: Update of dataset for formal run
July 3—July 28, 2023: Formal run

Late submission
August 1—August 31, 2023: Late submission

NTCIR-17 Conference
August 1, 2023: Release of the draft task overview paper
September 1, 2023: Submission deadline for participant papers
November 1, 2023: Deadline for camera-ready versions of participant papers
December 12—December 15, 2023: NTCIR-17 Conference

6 PARTICIPATION

Ten teams registered for the task, but only six teams participated actively, i.e., submitted results for the formal run. Table 3 lists the active participating teams.

Table 3: Active participating teams

Group ID	Organization
KSU	Kyoto Sangyo University
fuys*	Fukuoka University
OUC*	Otaru University of Commerce
FA	Fast Accounting Co., Ltd.
JPXIteam	JPX Market Innovation & Research, Inc.
tkl2023†	Osaka Electro-Communication University
TO*	task organizers (baseline system)

*Task organizer(s) are in team
†No submissions for formal run (only late submissions)

Table 4 shows number of submissions in the formal run.

The characteristic aspects of the participating teams' systems and their contributions are listed in Tables 5 and 6, for the TDE and TTRE subtasks, respectively.

7 EXPERIMENTAL RESULTS

Tables 7 and 8 list the evaluation results of the TDE and TTRE subtasts in the formal run, respectively. Table 7 lists the accuracy scores of the TDE subtask in the formal run. For instance, Team KSU with ID 81 achieved an accuracy score of 0.9537, and Team FA with ID 100 achieved a score of 0.9343. Table 8 lists the system

Table 4: Number of submissions in formal run

Team	TDE	TTRE	Total
KSU	1	1	2
fuys*	-	22	22
OUC*	13	2	15
FA	21	1	22
JPXIteam	3	-	3
tkl2023	-	-	-
TO*	1	-	1
Total	39	26	65

performances on the basis of three metrics: "Name," "Value," and "Total" for the TTRE subtask during the formal run. For example, Team "fuys" with ID 95 achieved F1 scores of 0.2707 for "Name", 0.1943 for "Value", and an overall score of 0.2325 for "Total".

See Appendix for the results of dry run and the late submissions.

8 CONCLUSION

We organized the NTCIR-17 UFO task and in this paper introduced the task definition, dataset, and evaluation methodology of two subtasks. The UFO task attracted six research groups. In the TDE (table data extraction) subtask, there were 10 dry runs and 39 formal runs contributed. In the TTRE (text-to-table relationship extraction) subtask, 13 dry runs and 26 formal runs were contributed.

REFERENCES

[1] Najah-Imane Bentabet, Rémi Juge, Ismail El Maarouf, Virginie Mouilleron, Dialekti Valsamou-Stanislawski, and Mahmoud El-Haj. 2020. The Financial Document Structure Extraction Shared task (FinToc 2020). In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*. COLING, Barcelona, Spain (Online), 13–22. https://aclanthology.org/2020.fnp-1.2

[2] Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2020. Overview of the NTCIR-15 FinNum-2 Task: Numeral attachment in financial tweets. In *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo Japan*.

[3] Dominique Mariko, Hanna Abi Akl, Estelle Labidurie, Stéphane Durfort, Hugues de Mazancourt, and Mahmoud El-Haj. 2020. Financial Document Causality Detection Shared Task (FinCausal 2020). arXiv:2012.02505

Table 5: Summary of participants' systems for TDE subtask.

Team	Description	
FA	Overview	ELECTRA-based classification and post-correction using Levenshtein edit distance
	Architecture	ELECTRA
	Input	The cell content and the entire row contents
OUC	Overview	BERT-based classification
	Architecture	BERT with further pre-training using unlabeled securities reports
	Input	The cell content
KSU	Overview	Classification using tree-based Transformer (TUTA)
	Architecture	TUTA-implicit
	Input	The machine-translated cell content with several features
JPXIteam	Overview	Classification using ChatGPT
	Architecture	ChatGPT
	Input	The cell content and a few examples

Table 6: Summary of participants' systems for TTRE subtask.

Team	Description	
fuys	Overview	Exact matching, or BERT classification for Name; the same row/column for Value
	Architecture	BERT
	Input	The target phrase and each cell content
OUC	Overview	ChatGPT-based extraction from a table following the target phrase
	Architecture	ChatGPT
	Input	The target phrase and the entire table
KSU	Overview	Cosine similarity for Name; the TDE classifier for Value
	Architecture	Multilingual-E5
	Input	The target phrase and each cell content
FA	Overview	Edit distance for Name; rates of numeric characters for Value
	Architecture	Rule-based
	Input	The target phrase and each cell content
tkl2023	Overview	Text embedding based on table-to-markdown
	Architecture	llama2 with QLoRA tuning
	Input	Answers and table-to-markdown output from ChatGPT

Table 7: Accuracy scores of TDE subtask in formal run

ID	Team	Accuracy
81	KSU	0.9537
100	FA	0.9343
89	FA	0.9338
84	FA	0.9317
85	FA	0.9291
72	FA	0.9236
78	FA	0.9236
88	OUC	0.9217
64	OUC	0.9195
79	FA	0.9175
65	FA	0.9158
66	FA	0.9146
56	OUC	0.9145
63	FA	0.9140
83	FA	0.9138
67	OUC	0.9129
76	OUC	0.9117
42	OUC	0.9113
61	FA	0.9104
70	FA	0.9101
41	OUC	0.9088
82	OUC	0.9066
96	OUC	0.9036
68	OUC	0.9004
35	FA	0.8945
53	FA	0.8938
33	FA	0.8913
55	FA	0.8875
36	FA	0.8871
34	FA	0.8865
62	OUC	0.8798
94	JPXIteam	0.8287
93	JPXIteam	0.8287
29	TO (baseline)	0.7981
91	JPXIteam	0.7919
47	FA	0.5472
40	OUC	0.4657
39	OUC	0.4657
69	FA	0.4546

Table 8: F1 Scores of TTRE subtask in formal run

ID	Team	Name	Value	Total
95	fuys	0.2707	0.1943	0.2325
97	fuys	0.2707	0.192	0.2314
60	fuys	0.2677	0.1859	0.2268
90	fuys	0.2707	0.1804	0.2256
87	fuys	0.2707	0.1793	0.2250
73	fuys	0.2677	0.1818	0.2248
57	fuys	0.2642	0.1771	0.2206
75	fuys	0.2677	0.1717	0.2197
52	fuys	0.2677	0.1707	0.2192
59	fuys	0.2636	0.1685	0.2161
77	fuys	0.2707	0.1580	0.2144
45	fuys	0.2638	0.1624	0.2131
86	fuys	0.2707	0.1550	0.2129
46	fuys	0.2630	0.1623	0.2126
50	fuys	0.2600	0.1640	0.2120
58	fuys	0.2568	0.1648	0.2108
44	fuys	0.2574	0.1590	0.2082
43	fuys	0.2559	0.1577	0.2068
51	fuys	0.2488	0.1596	0.2042
38	fuys	0.2490	0.1585	0.2038
37	OUC	0.1447	0.1823	0.1635
74	fuys	0.2677	0.0507	0.1592
80	fuys	0.2707	0.0361	0.1534
98	OUC	0.1018	0.1146	0.1082
99	KSU	0.0918	0.0408	0.0663
49	FA	0.0341	0.0131	0.0236

A RESULTS OF DRY RUN

Table 9 shows the number of submissions in the dry run. Table 10 lists the accuracy scores of the TDE subtask in the dry run and Table 11 lists the F1 scores of the TTRE subtask in the dry run, respectively. We updated the TTRE dataset during the dry run, and the results are shown separately.

Table 9: Number of submissions in dry run

Team	TDE	TTRE		Total
		20230501	20230613	
KSU	-	-	-	-
fuys*	-	6	6	12
OUC*	7	-	-	7
FA	1	-	-	1
JPXIteam	-	-	-	-
TO*	1	1	-	2
Total	10	7	6	23

Table 10: Accuracy scores of TDE subtask in dry run

ID	Team	Accuracy
1	TO (baseline)	0.8482
24	OUC	0.7956
23	OUC	0.7914
12	OUC	0.7832
21	OUC	0.7739
14	OUC	0.6285
22	OUC	0.5720
20	OUC	0.5636
27	FA	0.4153

Table 11: F1 scores of TTRE subtask in dry run

ID	Team	Name	Value	Total
20230501				
10	fuys	0.2654	0.2109	0.2382
13	fuys	0.2605	0.1758	0.2182
8	fuys	0.2478	0.1794	0.2136
9	fuys	0.2351	0.1799	0.2075
7	fuys	0.1822	0.1537	0.1680
6	fuys	0.0115	0.0681	0.0398
4	TO (random)	0.0002	0.0004	0.0003
20230613				
18	fuys	0.2484	0.1613	0.2049
19	fuys	0.2486	0.1606	0.2046
25	fuys	0.2463	0.1609	0.2036
17	fuys	0.2438	0.1601	0.2019
15	fuys	0.2059	0.1379	0.1719
16	fuys	0.1843	0.1321	0.1582

B RESULTS OF LATE SUBMISSIONS

Table 12 shows the number of late submissions, Table 13 lists the accuracy scores of the TDE subtask for late submissions, and Table 14 lists the F1 scores of the TTRE subtask for late submissions.

Table 12: Number of late submissions

Team	TDE	TTRE	Total
KSU	6	39	45
fuys*	-	2	2
OUC*	1	-	1
FA	5	1	6
JPXIteam	-	-	-
tkl2023	-	1	1
TO*	-	-	-
Total	12	43	55

Table 13: Accuracy scores of TDE subtask for late submissions

ID	Team	Accuracy
151	KSU	0.9503
140	KSU	0.9487
149	KSU	0.9480
139	KSU	0.9460
141	KSU	0.9459
150	KSU	0.9455
148	KSU	0.9438
142	KSU	0.9422
109	FA	0.9321
101	FA	0.9290
102	FA	0.9257
103	OUC	0.9145
108	FA	0.8134
107	FA	0.7937

Table 14: F1 scores of TTRE subtask for late submissions

ID	Team	Name	Value	Total
127	KSU	0.3221	0.2719	0.2970
146	KSU	0.3221	0.2704	0.2962
128	KSU	0.3212	0.2679	0.2945
125	KSU	0.3198	0.2688	0.2943
126	KSU	0.3190	0.2694	0.2942
156	KSU	0.3221	0.2659	0.2940
147	KSU	0.3212	0.2662	0.2937
144	KSU	0.3198	0.2669	0.2934
145	KSU	0.3190	0.2677	0.2934
154	KSU	0.3212	0.2629	0.2920
124	KSU	0.3131	0.2703	0.2917
155	KSU	0.3190	0.2635	0.2913
158	KSU	0.3198	0.2625	0.2912
143	KSU	0.3131	0.2683	0.2907
157	KSU	0.3131	0.2645	0.2888
136	KSU	0.3000	0.2563	0.2781
152	KSU	0.3000	0.2545	0.2773
153	KSU	0.3000	0.2537	0.2768
129	KSU	0.2892	0.2422	0.2657
133	KSU	0.2810	0.2453	0.2632
137	KSU	0.2804	0.2450	0.2627
138	KSU	0.2805	0.2435	0.2620
131	KSU	0.2753	0.2422	0.2587
132	KSU	0.2753	0.2419	0.2586
130	KSU	0.2776	0.2356	0.2566
134	KSU	0.2739	0.2367	0.2553
162	fuys	0.2995	0.1857	0.2426
161	fuys	0.2707	0.1988	0.2348
122	KSU	0.3221	0.1186	0.2204
121	KSU	0.3190	0.1179	0.2185
123	KSU	0.3212	0.1153	0.2182
120	KSU	0.3198	0.1154	0.2176
119	KSU	0.3131	0.1183	0.2157
116	KSU	0.3000	0.1120	0.2060
117	KSU	0.2804	0.1078	0.1941
118	KSU	0.2805	0.1067	0.1936
113	KSU	0.2810	0.1051	0.1931
111	KSU	0.2753	0.1032	0.1892
112	KSU	0.2753	0.1029	0.1891
114	KSU	0.2739	0.1015	0.1877
110	KSU	0.1983	0.0700	0.1342
160	tkl2023	0.0502	0.0404	0.0453
115	FA	0.0008	0.0004	0.0006

C DATASET EXAMPLES

C.1 TDE

C.1.1 Input.

Listing 1: Example input for TDE subtask

```
<html>
...
<body>
  ...
  <table>
    <tr>
      <td ...
        data-ufo-tde-cell-id="S100ABCD-0101010-tab1-r1c1"
        data-ufo-tde-cell-type="header">...</td>
      <td ...
        data-ufo-tde-cell-id="S100ABCD-0101010-tab1-r1c2"
        data-ufo-tde-cell-type="header">...</td>
      ...
    </tr>
    <tr>
      <td rowspan="2" ...
        data-ufo-tde-cell-id="S100ABCD-0101010-tab1-r2c1"
        data-ufo-tde-cell-type="attribute">...</td>
      <td ...
        data-ufo-tde-cell-id="S100ABCD-0101010-tab1-r2c2"
        data-ufo-tde-cell-type="data">...</td>
      ...
    </tr>
    ...
  </table>
  ...
</body>
</html>
```

The `data-ufo-tde-cell-type` attribute is only provided in the training dataset.

C.1.2 Output.

Listing 2: Example output for TDE subtask

```
{
  ...
  "S100ABCD-0101010-tab1-r1c1": "header",
  "S100ABCD-0101010-tab1-r1c2": "header",
  "S100ABCD-0101010-tab1-r1c3": "header",
  "S100ABCD-0101010-tab1-r1c4": "header",
  "S100ABCD-0101010-tab1-r2c1": "attribute",
  "S100ABCD-0101010-tab1-r2c2": "data",
  "S100ABCD-0101010-tab1-r2c3": "data",
  "S100ABCD-0101010-tab1-r2c4": "data",
  "S100ABCD-0101010-tab1-r3c2": "data",
  "S100ABCD-0101010-tab1-r3c3": "data",
  "S100ABCD-0101010-tab1-r3c4": "data",
  ...
}
```

C.2 TTRE

C.2.1 Input.

Listing 3: Example input for TTRE subtask

```
<html>
...
<body>
  ...
  <p>
    <mark class="annotate"
      data-ttre-mark-id="S100ABCD-1-mark1"
      data-ttre-name-cell-ids="S100ABCD-1-tab0-r0c2
S100ABCD-1-tab0-r2c0"
      data-ttre-value-cell-ids="S100ABCD-1-tab0-r2c2"
      data-ttre-etc-cell-ids=""
    >The dividend per share of 80.00 yen for the fiscal
year ended March 31, 2016</mark> includes a commemorative
 dividend of 10.00 yen for the 60th anniversary of the
Company's founding.
  </p>
  ...
  <table data-ttre-table-id="S100ABCD-1-tab0">
    <tr>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r0c0">Period
</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r0c1">The 76
th (2015)</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r0c2">The 77
th (2016)</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r0c3">The 78
th (2017)</td>
    </tr>
    ...
    <tr>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r2c0">
Dividend per share (yen)</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r2c1
">60.00</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r2c2
">80.00</td>
      <td data-ttre-cell-id="S100ABCD-1-tab0-r2c3
">92.00</td>
    </tr>
    ...
  </table>
  ...
</body>
</html>
```

The `data-ttre-name-cell-ids`, `data-ttre-value-cell-ids` and `data-ttre-etc-cell-ids` attributes are only provided in the training dataset.

C.2.2 Output.

Listing 4: Example output for TTRE subtask

```
[
  ...
    {
    "S100ABCD-1-mark0": {
        "name": [],
        "value": []
    },
    "S100ABCD-1-mark1": {
        "name": [
            "S100ABCD-1-tab0-r0c2",
            "S100ABCD-1-tab0-r2c0"
        ],
        "value": [
            "S100ABCD-1-tab0-r2c2"
        ]
    },
  ...
  ...
]
```