

Overview of the NTCIR-18 U4

Unifying, Understanding, and Utilizing Unstructured Data in Financial Reports

Yasutomo Kimura

Otaru University of Commerce

Eisaku Sato

Otaru University of Commerce

Kazuma Kadowaki

The Japan Research Institute, Limited

Hokuto Ototake

Fukuoka University

- **Task:** Extracting and analyzing information from tables in annual securities reports — Table Retrieval and Table QA
- **Participants:** 10 active teams; 210 total submissions

Annual Securities Reports

- Help investors evaluate and select promising listed stocks
- 200–300 pages per report; approximately 221 tables

Tables in Annual Securities Reports

- Structured using XBRL: financial information (18%)
 - e.g., income statements, balance sheets, cash flow statements
- Unstructured tables: non-financial information
 - e.g., SDGs, corporate governance, business risks

Task Definitions

Table Retrieval

- Finding a table with the answer to a given question
- Input
 - Question (e.g., “株式会社バンダイナムコホールディングスの2020 年時点における「純資産額、経営指標等」は?”)
 - HTML files
- Output
 - Table ID (e.g., “S100ISF1-0101010-tab2”)

Table QA (Table Question Answering)

- Identifying the value that answers a given question
- Input
 - Question (e.g., “株式会社バンダイナムコホールディングスの2020 年時点における「純資産額、経営指標等」は?”)
 - Table ID (e.g., “S100ISF1-0101010-tab2”)
 - HTML file
- Output
 - Cell ID (e.g., “S100ISF1-0101010-tab2-r8c1”)
 - Value (e.g., “4448000000”, converted from “4,448 million yen”)

Dataset

Generated from annual securities reports of 289 companies, submitted between 2020-04-01 and 2021-03-31

Task	Train	Validation	Test	Total
Table Retrieval	10,299	1,441	2,897	14,637
Table QA	10,300	1,441	2,898	14,639

Schedule

Phase	Schedule
Dataset Release	2024-07-20
Dry Run	2024-07-20 - 10-31
Formal Run	2024-11-01 - 12-28
Results Return	2025-02-01
Conference	2025-06-10 - 06-13

Participants

Group ID	Organization	Submissions (Formal Run)		
		TR	TQA	Total
AKBL	Toyohashi Univ. Technol.	1	12	13
FUSINT*	Fukuoka Univ.	36	15	51
STMK24	Stockmark Inc.	-	9	9
tklb2024	Osaka Electro-Comm. Univ.	-	3	3
airev	AIREV, Inc.	22	11	33
SMM	Keio Univ.	5	6	11
UTUtLB25	Univ. Tsukuba	-	10	10
OUC*	Otaru Univ. Commerce	17	-	17
TMU2024	Taipei Med. Univ. & Shuang-Ho Hosp.	-	5	5
WhiteME	Mitsubishi Electr. Corp.	-	6	6
TO*	task organizers (baseline system)	3	2	5
Total		84	79	163

Table Retrieval Task

ID	Team	Public	Private
		Table ID	
62	SMM	0.9770	0.9755
63	SMM	0.9777	0.9718
150	airev	0.9117	0.9115
201	airev	0.9256	0.8952
189	FUSINT	0.8772	0.8808
181	FUSINT	0.8910	0.8689
108	OUC	0.5771	0.6117
88	AKBL	0.3569	0.3287
66	TO (baseline)	0.2134	0.1870

Two-stage retrieval using BERT variants

- Candidates → reranking/classification

Table QA Task

ID	Team	Public		Private	
		Cell ID	Value	Cell ID	Value
219	UTUtLB25	0.9309	0.9470	0.9066	0.9317
220	airev	0.9785	0.9493	0.9473	0.9047
210	STMK24	0.9685	0.9163	0.9530	0.8483
131	SMM	0.8634	0.8657	0.8276	0.8194
156	FUSINT	0.9271	0.8849	0.8752	0.8125
172	FUSINT	0.9210	0.8810	0.8828	0.8063
164	FUSINT	0.9271	0.8872	0.8752	0.8025
209	WhiteME	0.7882	0.7467	0.7392	0.6903
127	AKBL	0.7797	0.7168	0.7850	0.6871
128	AKBL	0.7836	0.7214	0.7824	0.6834
68	TO (baseline)	0.0000	0.6899	0.0000	0.6470
194	airev	0.9747	0.6861	0.9605	0.6370
212	tklb2024	0.0000	0.5464	0.0000	0.4959
213	tklb2024	0.0000	0.5733	0.0000	0.4940
90	TMU2024	0.0223	0.1903	0.0345	0.1774
78	TMU2024	0.0284	0.1466	0.0426	0.1549
53	TO (baseline)	0.0138	0.0107	0.0125	0.0207

Candidate rows/columns/cells → answer selection/generation

- Using question–table pairs; optionally structural info
- Supervised classifiers vs. zero-shot reasoning using prompt

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

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 “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**

Output A “value” that is the answer to the question