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/10.20736/0002001332

Overview of NTCIR-17

Takehiro Yamamoto
University of Hyogo
Japan
t.yamamoto@sis.u-hyogo.ac.jp

Zhicheng Dou
Renmin University of China
China
dou@ruc.edu.cn

ABSTRACT

This is an overview of NTCIR-17, the seventeenth sesquiannual research project for evaluating information access technologies. NTCIR-17 involved various evaluation tasks related to information retrieval, natural language processing, question answering, etc. Nine tasks were organized in NTCIR-17. This paper describes an outline of NTCIR-17, which includes its organization, schedule, scope, and task designs. In addition, we introduce brief statistics of the NTCIR-17 participants. Readers should refer to individual task overview papers for their detailed descriptions and findings.

1 INTRODUCTION

Since 1997, the NTCIR project has promoted research efforts for enhancing Information Access (IA) technologies such as Information Retrieval, Question Answering, and Natural Language Processing technologies. Its general purposes are to (1) Offer a research infrastructure that allows researchers to conduct a large-scale evaluation of IA technologies, (2) Form a research community in which findings from comparable experimental results are shared and exchanged, and (3) Develop evaluation methodologies and performance measures of IA technologies. Collaborative works in NTCIR have allowed us to create large-scale test collections that are indispensable for confirming the effectiveness of novel IA techniques. Moreover, in the collaboration process, it is expected that deep insight into research problems is successfully shared among researchers. The ongoing NTCIR-17 aims to benefit all researchers who wish to advance their research efforts. For the details and characteristics of what has been proposed in NTCIR, readers should refer to the book [11].

2 OUTLINE OF NTCIR-17

2.1 Organization

The project of NTCIR-17 was directed by General Co-Chairs (GCCs): Charles Clarke (University of Waterloo), Noriko Kando (National Institute of Informatics), Makoto P. Kato (Tsukuba University), and Yiqun Liu (Tsinghua University). Under the supervision of GCCs, Program Committee (PC) reviewed task proposals that were submitted according to a call for task proposals and made acceptance decisions for NTCIR-17. The members of the PC are Chung-Chi Chen (National Institute of Advanced Industrial Science and Technology), Hsin-Hsi Chen (National Taiwan University), Gareth Jones (Dublin City University), Noriko Kando (National Institute of Informatics), Makoto P. Kato (University of Tsukuba), Yiqun Liu (Tsinghua University), Alistair Moffat (The University of Melbourne), Jian-Yun Nie (University de Montreal), Douglas Oard (University of Maryland), Tetsuya Sakai (Waseda University), Mark Sanderson (RMIT University), and Ian Soboroff (NIST). After the review by PC, organizers of accepted tasks have promoted research activities of NTCIR-17 under the coordination of the two Program Co-Chairs (PCCs).

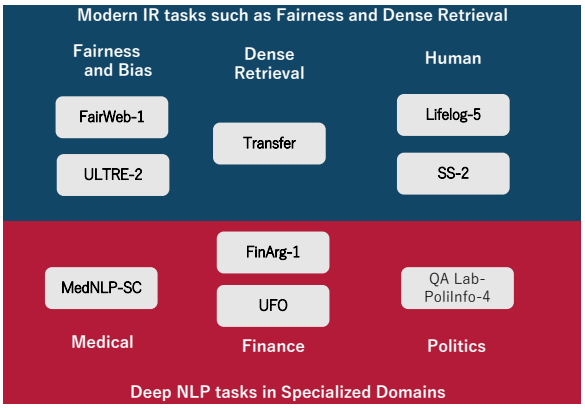

Figure 1: Overview of the NTCIR-17 tasks.

2.2 Schedule and Research Activities

A call for task proposals was released in May 2022, and 6 tasks (3 core tasks and 3 pilot tasks) were selected. To encourage more diverse tasks, a call for additional task proposals was released in September 2022, and three tasks (two core tasks and one pilot task) were selected. In total, five core tasks and four pilot tasks were organized in NTCIR-17. While Lifelog-5, SS-2, and ULTRE-2 were selected in the additional call for task proposals, other tasks were selected in the call for task proposals. Actual NTCIR-17 activities started in September 2022, and a kickoff event was also held in September 2022. According to the purpose and policy of each task, datasets for experiments (documents, queries, and so on) were developed by the task organizers, and distributed to participants (*i.e.*, research groups or teams participating in the task) by the organizers. New test collections were created based on the evaluation of results that were submitted by participants. The research outcome will be reported at the NTCIR-17 conference to be held at National Institute of Informatics, Tokyo from December 12th to 15th, 2023.

2.3 Scope and Tasks

The core task explores problems that have been known well in the fields of IA, while the pilot task aims to address novel problems for which there are uncertainties as to how to evaluate them. Figure 1 summarizes the 5 core tasks (FinArg-1, Lifelog-5, MedNLP-SC, QA Lab-PoliInfo-4, and SS-2) and 4 pilot tasks (FairWeb-1, Transfer, UFO, and ULTRE-2) organized in the NTCIR-17.

- Information Retrieval: Modern IR tasks

- Natural Language Processing: Deep language understanding in specialized domains such as finance, politics, and medical treatment.

It is interesting to see that the task covers a wide range of research topics in modern information retrieval such as fairness in IR (FairWeb-1), dense retrieval (Transfer), session-based information retrieval (SS-2), lifelog retrieval (Lifelog-4), and unbiased learning to rank (ULTRE-2). It is also worth mentioning that some tasks tackled deep NLP tasks in specialized domain such as financial documents (FinArg-1 and UFO), medical text (MedNLP-SC), and political documents (QA Lab-PoliInfo-4)

3 OUTLINE OF NTCIR-17 TASKS

NTCIR-17 has five core tasks and four pilot tasks. They cover information access problems in various domains, such as finance, medicine, social media, and politics. The overview of these tasks are summarized as follows.

3.1 FairWeb-1 (Pilot Task) [13]

FairWeb-1 represents an English web search task that departs from the traditional ad-hoc document retrieval model, such as the WWW tasks [12] in previous NTCIR conferences. This task, instead, includes a dual perspective, considering document relevance from the standpoint of search engine users and group fairness from the viewpoint of the entities being sought.

FairWeb-1 focuses on three distinct entity types: researchers (R), movies (M), and Youtube contents (Y). Search topics are carefully tailored to describe the information needs relevant to these entities. Each entity type is associated with one or two attribute sets, containing either nominal or ordinal groupings designed to ensure group fairness. Additionally, a target distribution is provided for each attribute set. In light of these elements, participants are asked to submit results that not only include relevant documents at the top rank but also exhibit group fairness in alignment with the attributes specified for each entity type.

3.2 FinArg-1 (Core Task) [2]

FinArg-1 represents an evolution from the preceding FinNum tasks [1], which exclusively centered on financial numerals, omitting the broader context of financial documents. Consequently, FinArg-1 introduces a novel facet of financial argument comprehension.

Specifically, FinArg-1 contains two tasks, namely argument-based sentiment analysis and argumentative relation identification with discussion threads. In the argument-based sentiment analysis task, participants are required to first classify the given sentences into “claim” or “premise”, and then further classify their sentiment into {bullish,bearish,neutral} or {positive,negative,neutral}. The second task consists of two distinct datasets, one in English (Earnings Call) and the other in Chinese (Social Media). Participants are asked to identify the argument units or relations in financial texts. These tasks will promote the study of fine-grained information embedded in the financial documents.

3.3 MedNLP-SC (Core Task) [8, 14]

MedNLP-SC combines several MedNLP tasks in previous NTCIR [15], while expanding its language scope. The primary objective is to foster the development of practical medical NLP applications within hospital settings.

MedNLP-SC has two subtasks: Social Media Adverse Drug Event Detection (SM-ADE) and Radio Report TNM staging (RR-TNM). SM-ADE is dedicated to the identification of adverse drug events (ADE), which involve discerning symptoms resulting from drug usage, within social media texts across multiple languages such as Japanese, English, French, and German. In contrast, RR-TNM evaluates the capability of NLP techniques to autonomously ascertain the clinical stage of lung cancer from radiology reports, necessitating clinical expertise and intricate reasoning. To facilitate experiments, a Japanese radiology report dataset has been meticulously anonymized and is made available.

3.4 Lifelog-5 (Core Task) [17]

LifeLog-5 is the fifth iteration of the LifeLog task. It is designed to improve the evaluation of various methods for both automatic and interactive information retrieval from multimodal lifelog archives. Distinguishing itself from its predecessor, LifeLog-4 in NTCIR-16 [18], LifeLog-5 leverages the LSC’23 dataset, which encompasses data obtained from an active lifelogger.

This task comprises three distinct subtasks: The Lifelog Semantic Access Task (LSAT), Lifelog Insight subTask (LIT), and Lifelog Question Answer subTask (LQAT). LSAT remains consistent with the previous iteration, focusing on retrieving specific moments (i.e., semantic events or activities) in the lifelogger’s life. LIT, on the other hand, lacks a formal evaluation process, emphasizing participants’ ability to offer insights into the lifelog data and devise effective data visualization techniques. Lastly, LQAT introduces an extended 85-day lifelog collection along with a substantial set of multiple-choice questions, prompting participants to develop systems capable of automatically responding to these questions.

3.5 QA Lab-PoliInfo-4 (Core Task) [10]

QA Lab-PoliInfo-4 builds upon the foundation of its predecessors, QA Lab-PoliInfo-1/2/3 [6], with a mission to advance the development of complex QA techniques in the context of Japanese political information, using materials like local assembly minutes and newsletters.

QA Lab-PoliInfo-4 has four primary subtasks: Question Answering-2, Answer Verification, Stance Classification-2, and Minutes-to-Budget Linking. The Question Answering-2 subtask, consistent with the one featured in NTCIR-16 QA Lab-PoliInfo-3, revolves around providing concise responses to questions based on assembly minutes. The Answer Verification subtask, acombination of Question Answering and Fact Verification subtasks from the previous iteration, QA Lab-PoliInfo-3, concentrates on validating the results generated by the Question Answering subtask, ensuring factual accuracy. Stance Classification-2, succeeding the Stance Classification subtask in NTCIR-15 QA Lab-PoliInfo-2, aims to deduce the stances of assembly members regarding bills from their speeches. Finally, the Minutes-to-Budget Linking subtask, inheriting aspects from the Budget Argument Mining in NTCIR-16 QA Lab-PoliInfo-3, seeks to identify argumentative components related to budget items and subsequently classify them based on their argumentative roles.

3.6 Transfer (Pilot Task) [4]

The Transfer Task is dedicated to the development of a comprehensive suite of technology for transferring resources originally generated for one purpose to another, within the context of dense retrieval applied to Japanese texts. It reuses the ad-hoc retrieval collections from NTCIR-1 and NTCIR-2, employing them as both training and evaluation datasets.

This task comprises two distinct subtasks: Dense First Stage Retrieval and Dense Reranking. The Dense First Stage Retrieval essentially mirrors an ad-hoc retrieval challenge. Participants are tasked with utilizing the title field of the original topic files as the query and are required to retrieve the top 1,000 documents based on their relevance to the query. The Dense Reranking subtask, on the other hand, is structured to facilitate the development of second-stage retrieval techniques within a multi-stage retrieval framework. Participants are presented with the top 1,000 documents retrieved by BM25 and are instructed to further rank them, offering an effective approach to optimizing the final ranking list.

3.7 SS-2 (Core Task) [7]

SS-2 maintains a similar framework to NTCIR-16 SS-1 [3] and aims at enhancing ranking strategies in context-aware search scenarios. In light of the ever-evolving landscape of search engines and the increasingly complex information needs of users, the conventional single-shot query scenario is gradually giving way to multiple interactions with search engines. In this case, user intent may undergo alterations, posing challenges to existing ad-hoc search methods.

Specifically, SS-2 comprises three primary subtasks: Fully Observed Session Search (FOSS), Partially Observed Session Search (POSS), and Session-level Search Effectiveness Estimation (SSEE). In line with the preceding iteration, the FOSS subtask offers complete contextual information leading up to the last query, whereas the POSS subtask provides only limited contextual data. Introducing a novel element, the SSEE task requires participants to leverage user feedback for the formulation of new session-level search effectiveness assessment metrics. This highlights the task's focus on adapting to the evolving nature of user interactions in search scenarios.

3.8 ULTRE-2 (Pilot Task) [9]

ULTRE-2 follows in the footsteps of NTCIR-16 ULTRE [16], striving to address the challenges associated with pseudo synthetic clicks and limited dataset size. It seeks to evaluate the effectiveness of ULTR (User-Level Task Retrieval) models by introducing a novel approach that harnesses a large-scale user behavior log sourced from a commercial web search engine.

In this context, ULTRE-2 constructs and releases a dataset founded on genuine user behavior logs obtained from Baidu, a prominent Chinese web search engine. The dataset includes initial ranking lists, query-document features, a wealth of user behavior data (such as clicks, displayed height, and displayed abstract), and display-related information (including dwelling time and slip count). Participants are tasked with training a feature-based ranking model on the provided training set and subsequently employing it to re-rank the ranking lists for the test queries. Furthermore, participants are encouraged to leverage the diverse forms of user behavior data and comprehensive display information to craft more sophisticated and effective ULTR models.

3.9 UFO (Pilot Task) [5]

The primary objective of the NTCIR-17 UFO task is to promote techniques for the extraction of structured information from tabular data and documents, with a particular focus on annual securities reports. These reports are typically authored in XBRL, an XML-based language. While they contain valuable information, the tables within these reports are characterized by non-uniform content scope and column organization.

To address these challenges, the UFO task introduces two key subtasks: table data extraction (TDE) and text-to-table relationship extraction (TTRE), both aimed at unraveling the structural intricacies of tables. TDE is centered on the precise extraction of entries and values within the tables present in annual securities reports. In contrast, TTRE seeks to establish connections between the values found in these tables and the corresponding statements within the textual content, enhancing the holistic understanding of the report's content.

4 PARTICIPANTS

Figure 2 shows the number of *active* (those who submitted results) participants. In the figure, the numbers are given for all the tasks from NTCIR-1 to NTCIR-17. At NTCIR-17, 49 research groups have participated in the tasks. The number of participants is almost the same as in the previous round. Note that some research groups participated in multiple tasks, which were counted as different groups. Readers should refer to the individual task overview papers for getting the picture of participants' approaches to each task. Also, they should refer to the participants' individual papers for detailed descriptions of their methods.

5 CONCLUSIONS

This paper described the overview of the seventeenth cycle of NTCIR carried from September 2022 to December 2023. NTCIR-17 has 9 evaluation tasks, which can be categorized into (1) modern information retrieval, and (2) natural language understanding in specialized domains. Most parts of the test collections developed by NTCIR-17 evaluation tasks will be released to non-participating research groups in the near future.

6 ACKNOWLEDGEMENTS

We would like to thank the organizers of all NTCIR-17 tasks for their tremendous amount of effort devoted to running successful tasks, the task participants for their valuable contributions to the IA research community, and the program committee members for their great suggestions for our accepted tasks. Finally, we would like to thank the current and past members of the NTCIR office for their continuous and careful support of our activity.

REFERENCES

[1] Chung-Chi Chen, Hen-Hsen Huang, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2022. Overview of the NTCIR-16 FinNum-3 Task: Investor's and Manager's Fine-grained Claim Detection. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 87–91.

Year	1999	2001	2002	2004	2005	2007	2008	2010	2011	2013	2014	2016	2017	2019	2020	2022	2023
Task/NTCIR round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total number	37	39	61	74	79	81	80	66	102	108	93	97	71	47	52	53	52
Automatic Term Recognition and Role Analysis (TMREC) (1)	9																
Ad hoc/Crosslingual IR (1) -> Chinese/English/Japanese IR (2) -> CLIR (3-6)	28	30	20	26	25	22											
Text Summarization Challenge (TSC) (2-4)		9	8	9													
Web Retrieval (WEB) (3-5)			7	11	7												
Question Answering Challenge (QAC) (3-6)			16	18	7	8											
Patent Retrieval [and Classification] (PATENT) (3-6)			10	10	13	12											
Multimodal Summarization for Trend Information (MUST) (5-7)					13	15	13										
Crosslingual Question Answering (CLQA) (5, 6) -> Advanced Crosslingual Information Access (ACLIA) (7, 8)					14	12	19	14									
Opinion (6) -> Multilingual Opinion Analysis (MOAT) (7, 8)						12	21	16									
Patent Mining (PAT-MN) (7, 8)							12	11									
Community Question Answering (CQA) (8)								4									
Geotemporal IR (GeoTime) (8, 9)								13	12								
Interactive Visual Exploration (Vis-Ex) (9)									4								
Patent Translation (PAT-MT)(7, 8) -> Patent Machine Translation (PatentMT)(9, 10)							15	8	21	21							
Crosslingual Link Discovery (Crosslink) (9, 10)									11	10							
INTENT(9, 10) -> Search Intent and Task Mining (IMine) (11, 12)									16	11	12	9					
One Click Access (1CLICK)(9, 10) -> Mobile Information Access (MobileClick) (11, 12)									4	8	4	11					
Recognizing Inference in Text (RITE)(9,10) -> Recognizing Inference in Text and Validation (RITE-VAL)(11)									24	28	23						
IR for Spoken Documents (SpokenDoc) (9, 10) -> Spoken Query and Spoken Document Retrieval (SpokenQuery&Doc) (11, 12)									10	12	11	7					
Mathematical Information Access (Math) (10, 11) -> MathIR (12)										6	8	6					
Medical Natural Language Processing (MedNLP) (10, 11) -> MedNLPDoc (12) -> MedWeb (13) -> Real-MedNLP(16) -> MedNLP-SC(17)										12	12	8		9		10	11
QA Lab for Entrance Exam (QALab) (11, 12, 13) -> QA Lab for Political Information (QALab-PoliInfo) (14, 15, 16, 17)											11	12	11	13	14	12	8
Temporal Information Access (Temporalia) (11, 12)											8	14					
Cooking Recipe Search (RecipeSearch) (11)											4						
Personal Lifelog Organisation & Retrieval (Lifelog) (12, 13, 14, 16, 17)												8	4	6		3	5
Short Text Conversation (STC) (12, 13, 14)												22	27	13			
Open Live Test for Question Retrieval (OpenLiveQ) (13, 14)													7	4			
Actionable Knowledge Graph (AKG) (13)													3				
Emotion Cause Analysis (ECA) (13)													3				
Neurally Augmented Image Labelling Strategies (NAILS) (13)													2				
We Want Web (WWW) (13, 14) -> We Want Web with CENTER (WWW) (15, 16)													5	4	8	3	
Fine-Grained Numeral Understanding in Financial Tweet (FinNum) (14,15,16)-> Fine-grained Argument Understanding in Financial Analysis (FinArg) (17)														6	7	7	11
CLEF/NTCIR/TREC REproducibility (CENTRE) (14)														1			
Dialogue Evaluation (DialEval) (15, 16)															7	4	
SHINRA 2020 Multi-lingual (SHINRA 2020-ML) (15)															7		
Data Search (Data Search) (15, 16)															5	6	
Micro Activity Retrieval Task (MART) (15)															5		
Session Search (SS) (16, 17)																3	2
Reading Comprehension for Information Retrieval (RCIR) (16)																3	
Unbiased Learning to Ranking Evaluation Task (ULTRE) (16, 17)																2	1
FairWeb (17)																	5
Resource Transfer Based Dense Retrieval (Transfer) (17)																	3
Understanding of non-Financial Objects in Financial Reports (UFO) (17)																	6

Figure 2: Number of active participants (from NTCIR-1 to NTCIR-17).

[2] Chung-Chi Chen, Chin-Yi Lin, Chr-Jr Chiu, Hen-Hsen Huang, Alaa Alhamzeh, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2023. Overview of the NTCIR-17 FinArg-1 Task: Fine-Grained Argument Understanding in Financial Analysis. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001323

[3] Jia Chen, Weihao Wu, Jiaxin Mao, Beining Wang, Fan Zhang, and Yiqun Liu. 2022. Overview of the NTCIR-16 Session Search (SS) Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 353–358.

[4] Hideo Joho, Atsushi Keyaki, and Yuki Oba. 2023. Overview of the NTCIR-17 Transfer Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001319

[5] Yasutomo Kimura, Hokuto Ototake, Kazuma Kadowaki, Takahito Kondo, and Makoto P. Kato. 2023. Overview of the NTCIR-17 UFO Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001321

[6] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Kazuma Kadowaki, Masaharu Yoshioka, Tomoyosi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Ken-Ichi Yokote, Tatsunori Mori, Kenji Araki, Teruko Mitamura, and Satoshi Sekine. 2022. Overview of the NTCIR-16 QA Lab-PoliInfo-3 Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 156–174.

[7] Haitao Li, Jia Chen, Jiannan Wang, Weihang Su, Qingyao Ai, Xinyan Han, Beining Wang, and Yiqun Liu. 2023. Overview of the NTCIR-17 Session Search (SS-2) Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001324

[8] Yuta Nakamura, Shouhei Hanaoka, Shuntaro Yada, Shoko Wakamiya, and Eiji Aramaki. 2023. NTCIR-17 MedNLP-SC Radiology Report Subtask Overview: Dataset and Solutions for Automated Lung Cancer Staging. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001328

[9] Zechun Niu, Jiaxin Mao, Qingyao Ai, Lixin Zou, Shuaiqiang Wang, and Dawei Yin. 2023. Overview of the NTCIR-17 Unbiased Learning to Rank Evaluation 2 (ULTRE-2) Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001320

[10] Yasuhiro Ogawa, Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Kazuma Kadowaki, Tomoyoshi Akiba, Minoru Sasaki, and Akio Kobayashi. 2023. Overview of the NTCIR-17 QA Lab-PoliInfo-4 Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001326

[11] Tetsuya Sakai, Douglas W Oard, and Noriko Kando. 2021. *Evaluating Information Retrieval and Access Tasks: NTCIR's Legacy of Research Impact*. Springer Nature.

[12] Tetsuya Sakai, Sijie Tao, Zhumin Chu, Maria Maistro, Yujing Li, Nuo Chen, Nicola Ferro, Junjie Wang, Ian Soboroff, and Yiqun Liu. 2022. Overview of the NTCIR-16 WeWantWeb with CENTRE (WWW-4) Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 231–242.

[13] Sijie Tao, Nuo Chen, Tetsuya Sakai, Zhumin Chu, Hiromi Arai, Ian Soboroff, Nicola Ferro, and Maria Maistro. 2023. Overview of the NTCIR-17 FairWeb-1 Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001318

[14] Shoko Wakamiya, Lis Kanashiro Pereira, Lisa Raithel, Katherine Yeh, Peitao Han, Seiji Shimizu, Tomohiro Nishiyama, Gabriel Herman Bernardim Andrade, Noriki Nishida, Hiroki Teranishi, Narumi Tokunaga, Philippe Thomas, Roland Roller, Pierre Zweigenbaum, Yuji Matsumoto, Akiko Aizawa, Sebastian Möller, Cyril Grouin, Thomas Lavergne, Aurélie Névéol, Patrick Paroubek, Shuntaro Yada, and Eiji Aramaki. 2023. NTCIR-17 MedNLP-SC Social Media Adverse Drug Event Detection: Subtask Overview. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001327

[15] Shuntaro Yada, Yuta Nakamura, Shoko Wakamiya, and Eiji Aramaki. 2022. Real-MedNLP: Overview of REAL document-based MEDical Natural Language Processing Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 285–296.

[16] Yurou Zhao, Zechun Niu, Feng Wang, Jiaxin Mao, Qingyao Ai, Yang Tao, Junqi Zhang, and Yiqun Liu. 2022. Overview of the NTCIR-16 Unbiased Learning to Rank Evaluation (ULTRE) Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 370–376.

[17] Liting Zhou. 2023. Overview of the NTCIR-17 Lifelog-5 Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. https://doi.org/10.20736/0002001329

[18] Liting Zhou, Cathal Gurrin, Graham Healy, Hideo Joho, Thanh-Binh Nguyen, Rami Albatal, and Frank Hopfgartner. 2022. Overview of the NTCIR-16 Lifelog-4 Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. 130–135.