Overview of NTCIR-16

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-16, the sixteenth sesquianual research project for evaluating information access technologies. NTCIR-16 involved various evaluation tasks related to information retrieval, natural language processing, question answering, etc. 10 tasks were organized in NTCIR-16. This paper describes an outline of NTCIR-16, which includes its organization, schedule, scope, and task designs. In addition, we introduce brief statistics of the NTCIR-16 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-16 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 [7].

2 OUTLINE OF NTCIR-16

2.1 Organization
The project of NTCIR-16 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) reviews task proposals that were submitted according to a call for proposal and made acceptance decisions for NTCIR-16. The members of the PC are Ben Carterette (Spotify) Hsin-Hsi Chen (National Taiwan University), Nicola Ferro (University of Padova), Gareth Jones (Dublin City University), Yiqun Liu (Tsinghua University), 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-16 under the coordination of the two Program Co-Chairs (PCCs).

2.2 Schedule and Research Activities
A call for task proposals was released in October 2020, and 8 tasks (5 core tasks and 3 pilot tasks) were selected. To encourage more diverse tasks, a call for additional task proposals was released in December 2022, and two tasks (one core task and one pilot task) were selected. In total, 6 core tasks and four pilot tasks were organized in NTCIR-16. While Lifelog-4 and RCIR were selected in the additional call for task proposals, other tasks were selected in the call for task proposals. Actual NTCIR-16 activities started in January 2021, and a kickoff event was held in March 2021. 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-16 conference to be held online from June 14th to 17th, 2022.

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 6 core tasks (Data Search 2, DialEval-2, FinNum-3, Lifelog-4, QA Lab-PoliInfo-3, and WWW-4) and 4 pilot tasks (RCIR, Real-MedNLP, SS, and ULTRE) organized in the NTCIR-16.

- Information Retrieval: Modern IR tasks from data to human.
- 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 information retrieval such as ad-hoc retrieval (WWW-4), session search (SS), data retrieval (Data Search 2), lifelog retrieval, and (3) Develop evaluation methodologies and performance measures of IA technologies.
Weibo, is used as the training and development set. Since the Chi-
As the open data movement increases, the need for a technique
The FinNum-1 task started at NTCIR-14 aiming to better un-
(English subtask) and data.gov (for English subtask) as in the
understand the claims in financial documents. The numerals often
play an important role in a deep understanding of the claims. For
example, the claim “the sales growth rate may exceed 40%.” makes
a stronger estimation than “the sales growth rate may exceed 40%.”
Understanding the numerals in such claims gives us a fine-grained
understanding of financial documents.

3.3 FinNum-3 (Core Task) [1]
The FinNum-1 task started at NTCIR-14 aiming to better under-
stand the numerals in financial documents. FinNum-3, which is
the successor of the FinNum-1 and FinNum-2 tasks, aims to un-
derstand the claims in financial documents. The numerals often
play an important role in a deep understanding of the claims. For
example, the claim “the sales growth rate may exceed 40%.” makes
a stronger estimation than “the sales growth rate may exceed 40%.”
Understanding the numerals in such claims gives us a fine-grained
understanding of financial documents.
FinNum-3 organized the claim detection task, in which a system
is asked to identify whether the given numeral is an in-claim or
out-of-claim. The system is also asked to classify the relevant cat-
egory for the numeral. The reports written by professional stock
analysts are used for Chinese subtask while the transcriptions of
companies’ earnings conference calls are used for English subtask.
The participants’ approaches to the task are diverse such as data
augmentation, numerical representation, knowledge-based, and
traditional machine learning.

3.4 Lifelog-4 (Core Task) [12]
Lifelog-4 is the successor of the LifeLog-1, LifeLog-2, and LifeLog-3
tasks which are organized in the NTCIR-12, 13, and 14, respectively.
The aim of the Lifelog task is to foster comparative benchmarking of
approaches to automatic and interactive information retrieval
from multimodal lifelog archives. One of the characteristics of the
Lifelog task is its dataset. Lifelog-4 uses the LSC’20 dataset, which
contains four months of lifelog data from one active lifelogger.
The dataset comprises (1) metadata such as time, location, and
biometrics, (2) images recorded by the wearable camera, and (3)
concepts annotated to these images.
Lifelog-4 organized one subtask called the Lifelog Semantic Ac-
cess (LEST) subtask, which is similar to the traditional ad-hoc
document retrieval. Given a topic, the system is required to retrieve
relevant images in the dataset. The topics contain ad-hoc topics
and know-item topics. An example ad-hoc topic is “find examples
of when was looking inside the refrigerator at home.” The system
is allowed to be either automatic or interactive. The interactive
system allows a user actively interact with the system while the
automatic system allows no interaction from the user except for
query formulation.

3.5 QA Lab-PoliInfo-3 (Core Task) [6]
QA Lab-PoliInfo aims to explore the techniques for real-world
complex question answering tasks. The spread of fake news is
becoming a critical social problem. Precise understanding of the
facts and opinions in political documents, which can be seen as a
primary source of political information, are important to combat
fake news.
QA Lab-PoliInfo-3 task is the third round of the QA Lab-PoliInfo
task, which started at the NTCIR-14. Taking over the success of
QA Lab-PoliInfo-1 and QA Lab-PoliInfo-2, the QA LabPoliInfo-3
task organized a variety of subtasks, namely, QA Alignment, Que-
sion Answering, Fact Verification, and Budget Argument Mining
subtasks. Diverse political documents are also provided to the par-
ticipants, such as the minutes of the Tokyo Metropolitan Assembly,
newsletters of the Tokyo Metropolitan Government, budget infor-
mation of the National Diet, and several prefectures and cities.

3.6 WWW-4 (Core Task) [8]
The We Want Web with CENTRE (WWW-4) Task is the fourth
round of the WWW task series which aims to evaluate the effective-
ness of adhoc web search algorithms. Ranking is a core component
of web search engines, and it has been studied for decades. In recent
years, many new neural retrieval algorithms are proposed and it
would be interesting to quantify the technical improvements of
the recent approaches. WWW-4 focuses on the adhoc English web
search task, and it keeps the same requirement of previous tasks:
given a query set and a document corpus, returning top ranked
documents from the corpus for each query.
In WWW-4, there are two main changes. Firstly, a new English
web corpus, namely Chuweb21, is introduced. Chuweb21 is a subset
of the Common Crawl dataset and it contains 3,402,457 domains
and 858,616,203 web pages. Secondly, two versions of relevance
assessment are introduced: the Gold version given by the topic
creators, and the Bronze version labeled by “normal” assessors who
are neither topic creators nor topic experts.

3.7 RCIR (Pilot Task) [4]
RCIR is a pilot task that aims to understand the incorporation of
reading comprehension measures and eye tracker signals into the
process of document ranking. RCIR consists of two subtasks: a
comprehension-evaluation task (CET) and a comprehension-based
retrieval task (CRT). The former aims to predict a person’s com-
prehension level by exploiting eye movement information when
reading a passage, and the latter aims to explore the methods of
integrating comprehension evidence into passage retrieval systems.
RCIR creates a dataset by collecting the eye movements of ex-
perimental participants during their reading tasks with different
constraints and manipulations, and the corresponding answers of
the multiple-choice questions presented to experimental partici-
pants. The questions are used to measure the comprehension level
of the participants.
In terms of evaluation measures, RCIR uses Spearman’s corre-
lation coefficient for the CET subtask, and uses Normalized Dis-
counted Cumulative Gain for the CRT subtask.

3.8 Real-MedNLP (Pilot Task) [10]
The pilot task Real-MedNLP is designed to explore the natural
language processing techniques in medical fields. It is the succe-
sor of the four previous MedNLP tasks: MedNLP-1, MedNLP-2,
MedNLPDoc, and MedWeb. Different from these previous tasks,
Real-MedNLP introduces real clinical text datasets: the MedTxt-CR
corpus containing case reports and the MedTxt-RR corpus contain-
ing radiology reports. The original datasets are in Japanese, and
are translated into English.
With the support of the real clinical text corpus, Real-MedNLP
offers subtasks on few-resource named entity recognition, and
adverse drug event extraction.

3.9 SS (Pilot Task) [3]
SS is a pilot task aiming at exploring good ranking models for
text-aware search (i.e., session search). Existing adhoc search
models assume each query submitted to a search engine is stan-
dalone. However, in a real search scenario, a user may issue multiple
queries to a search system within a short time interval, to find the
information they need. Utilizing the contextual information, such as
the preceding queries and their clicks, has been proved beneficial
for generating better ranking results for the current query.
SS consists of two subtasks, namely the Fully Observed Session
Search (FOSS) task and the Partially Observed Session Search (POSS)
task. SS uses the TianGong-ST dataset [2] for training, and merges
the TianGong-SS-FSD and TianGong-Qref datasets for testing. The
over 100k training sessions in TianGong-ST are sampled from real
web search sessions from query logs of the Sogou search engine.
Among these sessions, 2,000 are manually assessed by humans. Dif-
differently, the test sessions are extracted from field studies conducted
by users.

3.10 ULTRE (Pilot Task) [11]
The ULTRE task is motivated by the advances in the trending re-
search topic “Unbiased Learning to Rank” which aims to learn a
stable ranking model from the noisy and biased user behaviour
data. It consists of two subtasks: the offline ULTR subtask and the
online ULTR subtask.
ULTRE constructs a dataset constructed based on SogouSRR. The
dataset includes 1,200 queries sampled from Sogou.com and HTML
sources of their top 10 search results. ULTRE uses real click logs
to train and calibrate click models. These models are then used to
generate synthetic user clicks for training queries for both subtasks.
Human relevance labels are used to evaluate the performance over
the test queries.

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-16. At NTCIR-16, 53 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 sixteenth cycle of NTCIR
carried from January 202 to June 2022. NTCIR-16 has 10 evaluation
tasks, which can be categorized into (1) traditional and novel in-
formation retrieval evaluation problems, and (2) natural language
understanding in specialized domains. Most parts of the test collec-
tions developed by NTCIR-16 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-16 tasks for
their tremendous amount of efforts devoted to running successful
tasks, the task participants for their valuable contributions to the
IA research community, and 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 NTCIR-16
Conference.
ST: A New Dataset with Large-scale Refined Real-world Web Search Sessions.
In Proceedings of the 28th ACM International Conference on Information and
Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019, Wenwu


### Figure 2: Number of active participants (from NTCIR-1 to NTCIR-16).

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