NTCIR

Overview of NTCIR-16

Takehiro Yamamoto   Zhicheng Dou

University of Hyogo   Renmin University of China

- **What is Evaluation Forum?**

- **Overview of NTCIR-16**

- **Introduction of NTCIR-16 Tasks**

- **Highlights of NTCIR-16 Conference**

- **Summary**

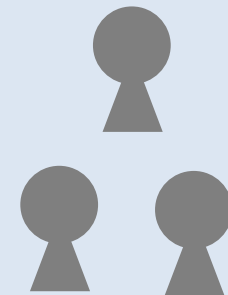- **Research groups gather together and tackle the same problems**
  (problem definition, test collection, evaluation measure, etc.)
- **The systems can be compared across participants**

collaboration

Participants
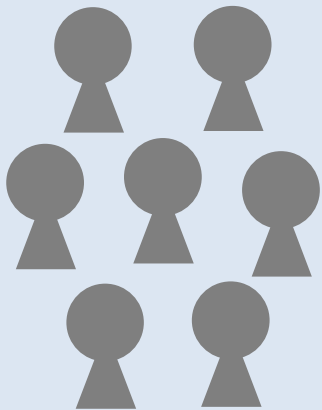
Task Organizers

Task and test collection

System output（Run）
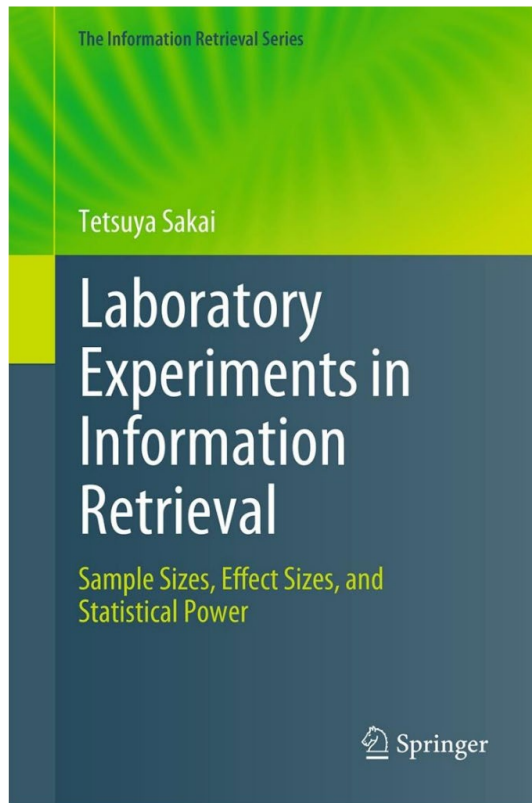
Evaluation Results

Share results , exchange ideas

- Can develop a large-scale test collections

- Can fairly compare the systems

- Can communicate with people
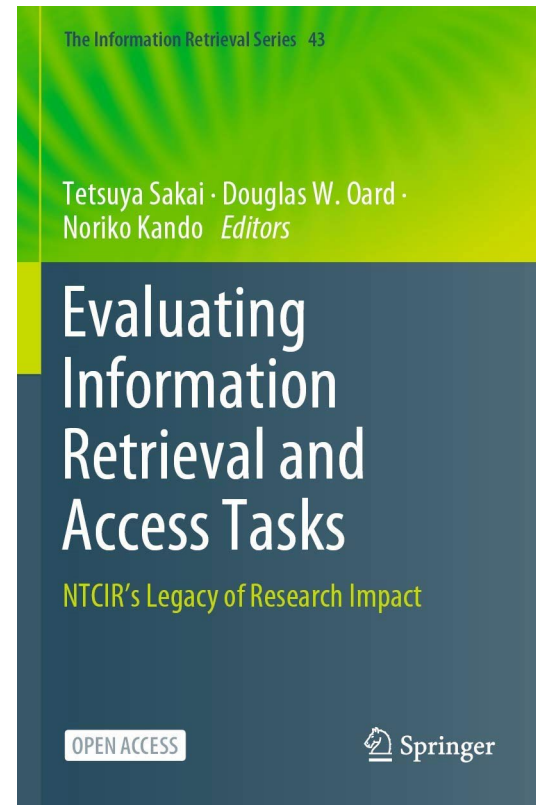 who has the same research interest

NTCIR

NII **T**estbeds and **C**ommunity for **I**nformation access **R**esearch

Recommended Books

Tetsuya Sakai: Laboratory Experiments in Information Retrieval: Sample Sizes, Effect Sizes, and Statistical Power (The Information Retrieval Series), Springer, 2018.

Tetsuya Sakai, Douglas W. Oard, Noriko Kando: Evaluating Information Retrieval and Access Tasks: NTCIR's Legacy of Research Impact (The Information Retrieval Series), Springer, 2021.

open access!

Overview of NTCIR-16
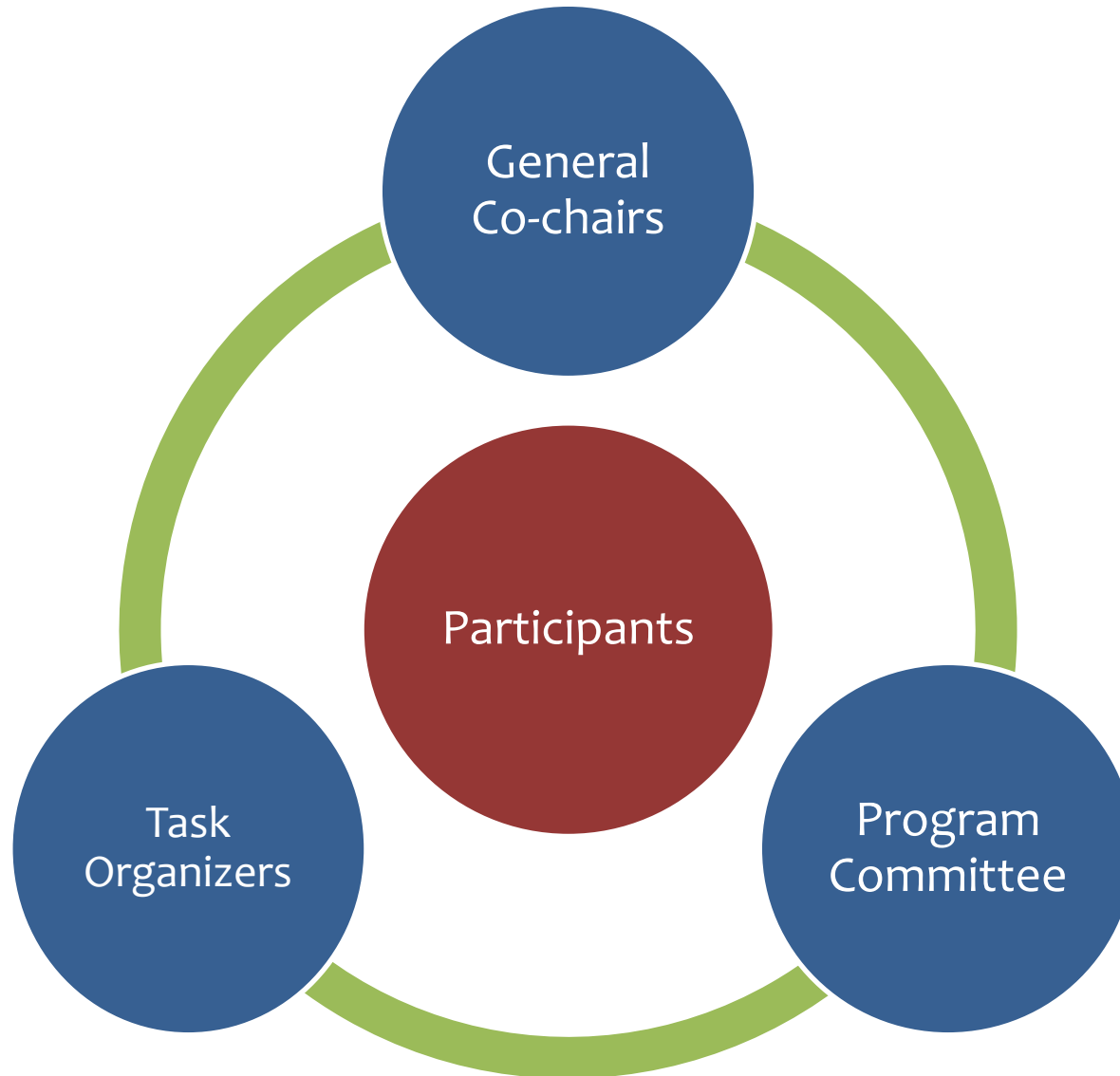
Noriko Kando

National Institute of Informatics, Japan

Charles L. A. Clarke

University of Waterloo, Canada

Makoto P. Kato

University of Tsukuba, Japan

Yiqun Liu

Tsinghua University, China

NTCIR-16 Program Committee (PC)

- Takehiro Yamamoto, University of Hyogo
- Zhicheng Dou, Renmin University of China
- Ben Carterette, Spotify
- Hsin-Hsi Chen, National Taiwan University
- Nicola Ferro, University of Padova
- Gareth Jones, Dublin City University
- Noriko Kando, NII
- Makoto P. Kato, University of Tsukuba
- Yiqun Liu, Tsinghua University
- Jian-Yun Nie, University of Montreal
- Douglas Oard, University of Maryland
- Tetsuya Sakai, Waseda University
- Mark Sanderson, RMIT University
- Ian Soboroff, NIST

Review Process

- **Call for Task Proposal**
  - Proposal due: Dec 4, 2020
  - **5 Core Tasks**, **3 Pilot Tasks** were accepted

- **Additional Call for Task Proposal**
  - Proposal due: Jan 15, 2021
  - **1 Core Task**, **1 Pilot Task** were accepted
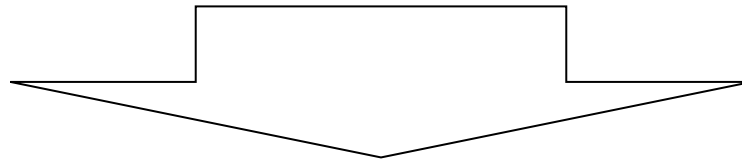
10 Tasks (6 Core tasks, 4 Pilot tasks) accepted

Two Task Types

- **Core Task**
  - To foster researches on a particular information access problem by providing researchers with a common ground for evaluation.

- **Pilot Task**
  - To focus on a novel information access problem and there are uncertainties either in task designing or organization.

Tasks in NTCIR-16

NTCIR-16 Tasks

CORE TASKS

1. **Data Search 2**: Data Search

2. **DialEval-2**: Dialogue Evaluation

3. **FinNum-3**: Investor's and Manager's Fine-grained Claim Detection

4. **Lifelog-4**: Lifelog Access and Retrieval

5. **QA Lab-PoliInfo-3**: Question Answering Lab for Political Information

6. **WWW-4**: We Want Web 4 with CENTRE

PILOT TASKS

7. **RCIR**: Reading Comprehension for Information Retrieval

8. **Real-MedNLP**: Real document-based Medical Natural Language Processing

9. **SS**: Session Search

10. **ULTRE**: Unbiased Learning to Ranking Evaluation Task

Ad-hoc retrieval for governmental **statistical data**
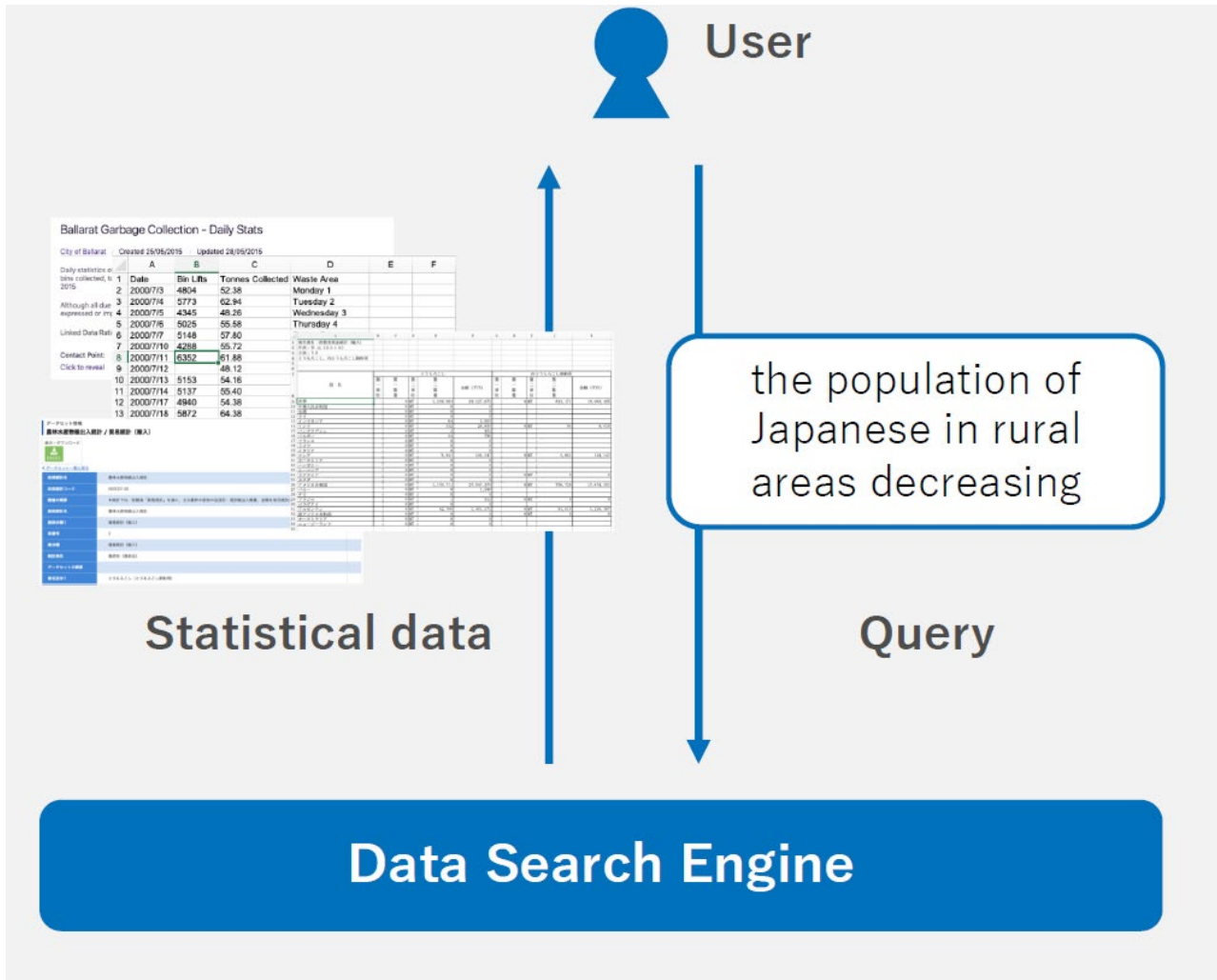
Subtasks:
- Ad-hoc retrieval
- Question Answering
- User interface

Figure from Data Search 2: NTCIR-16 Kick-Off Event, 2021.

Evaluation of the quality of a customer-helpdesk **dialogue**

Dialogue Quality Subtask

Nugget Detection Subtask

Chinese/English customer-helpdesk dialogue

Dialogue
- Customer's turn
- Helpdesk's turn
- Customer's turn
- Helpdesk's turn }Turn

Gold distribution for Task Accomplishment, Customer Satisfaction, Dialogue Effectiveness scores

Gold distribution over Regular nugget, Goal nugget, Not-A-Nugget

A nugget is a turn that helps the customer transition towards "Problem Solved"

Nugget types

Participant's estimated distribution

Evaluating distributions over ordinal bins

Evaluating distributions over nominal bins

Participant's estimated distribution

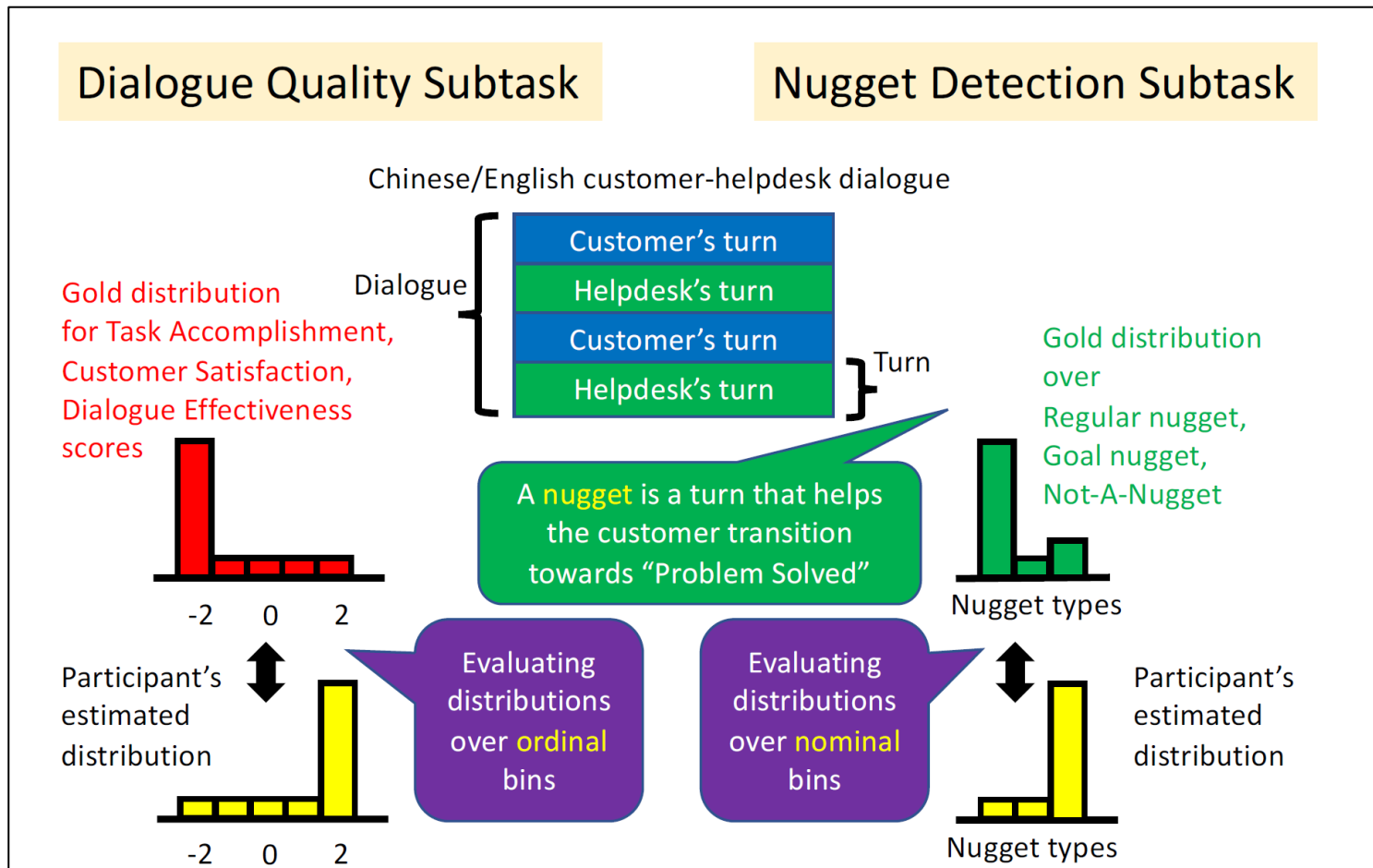Nugget types

Figure from DialEval-2: NTCIR-16 Kick-Off Event, 2021.

Fine-grained understanding of claims in **financial documents**

Investors always make a claim with an estimation

We estimate that the sales growth rate may exceed **80%**

- Detect the given numeral is in-claim or out-of-claim
- Classify the category of the numeral
- Documents: Investors reports, earnings conference calls

Automatic and interactive information retrieval from multimodal lifelog archive

Example topic
What's in the refrigerator?

Figure from Overview of Lifelig-4,
Proceedings of the NTCIR-16 Conference, 2022

- **LSC'21 dataset**
  - PoV cameras, smartwatches location, etc
  - 4 months of lifelog data

Aims to explore the techniques for real-world complex question answering tasks in **political information**

Figure from QA Lab-PoliInfo-3: NTCIR-16 Kick-Off Event, 2021.

Aims to quantify the progress and reproducibility of web search algorithms in offline ad hoc retrieval settings

Run Types
- **REV** (Revived) runs:
    - Top performer at WWW-3
- **New** Runs
- **REP** runs
    - Runs reproduced by another team

Corpus: **Chuweb21**
- Sampled from Common Crawl

Aims to understand the incorporation of reading comprehension measures and eye tracker signals into the process of document ranking.

Collect eye movements
of users while reading passages

Subtasks
- **Comprehension-evaluation task (CET)**
  - Predict a person's reading comprehension level from eye movement information
- **Comprehension-based retrieval task (CRT)**
  - Retrieve passages by integrating the comprehension evidence

(Real document-based Medical Natural Language Processing)

The goal of Real-MedNLP:

- Real datasets

- Bilingual capability

- Practicality

Subtasks:

- Few-resource Named Entity Recognition

- Adverse drug events extraction

Sample case report

Sample radiology report

Figures from Real-MedNLP: Overview of REAL document-based MEDical Natural Language Processing Task, Proceedings of the NTCIR-16 Conference, 2022

Aiming at exploring good ranking models for session search
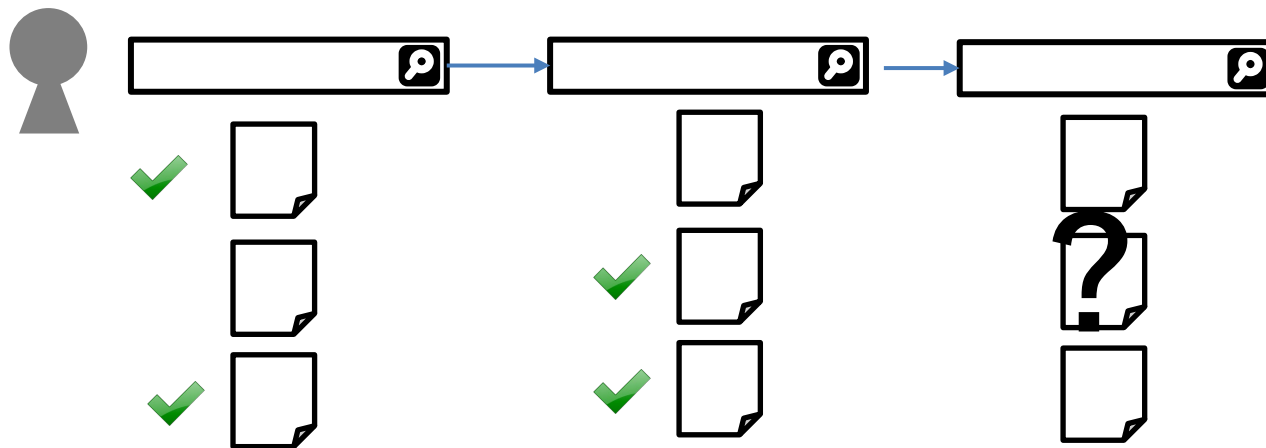
Dataset: **TianGong-ST**

- Sampled from
   real Web search logs

Subtasks:

- Fully observed

- Partially observed

- **Unbiased Learning to Rank** (**ULTR**)

  - aims to learn a stable ranking model from the noisy and biased user behavior data.

- How to properly evaluate and compare different ULTR approaches has not been systematically investigated

- Dataset is constructed from SogouSRR

- Click models are generated from the real user click logs

NTCIR-16 Tasks

CORE TASKS

1. **Data Search 2**: Data Search

2. **DialEval-2**: Dialogue Evaluation

3. **FinNum-3**: Investor's and Manager's Fine-grained Claim Detection

4. **Lifelog-4**: Lifelog Access and Retrieval

5. **QA Lab-PoliInfo-3**: Question Answering Lab for Political Information

6. **WWW-4**: We Want Web 4 with CENTRE

PILOT TASKS

7. **RCIR**: Reading Comprehension for Information Retrieval

8. **Real-MedNLP**: Real document-based Medical Natural Language Processing

9. **SS**: Session Search

10. **ULTRE**: Unbiased Learning to Ranking Evaluation Task

Languages covered by Tasks

Tasks	Chinese	Japanese	English
Data Search 2		✔	✔
DialEval-2	✔		✔
FinNum-3	✔		✔
Lifelog-4			✔
QA Lab-PoliInfo-3		✔	
WWW-4			✔
RCIR			✔
Real Med-NLP		✔	✔
SS	✔		
ULTRE	✔		

Task	Number of Active Participants
Data Search 2	6
DialEval-2	4
FinNum-3	7
Lifelog-4	3
QA Lab-PoliInfo-3	12
WWW-4	3
RCIR	3
Real Med-NLP	10
SS	3
ULTRE	2
Total # of unique active participants	**49**

Active participants: Research groups submitted final results for evaluation

One group can submit multiple systems (runs)

History of NTCIR Tasks

Year	1999	2001	2002	2004	2005	2007	2008	2010	2011	2013	2014	2016	2017	2019	2020	2022
Task/NTCIR round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Total number	37	39	61	74	79	81	80	66	102	108	93	97	71	47	52	53
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)			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)										12	12	8	9			10
QA Lab for Entrance Exam (QALab) (11, 12, 13) -> QA Lab for Political Information (QALab-PoliInfo) (14, 15, 16)											11	12	11	13	14	12
Temporal Information Access (Temporalia)(11, 12)											8	14				
Cooking Recipe Search (RecipeSearch) (11)											4					
Personal Lifelog Organisation & Retrieval (Lifelog) (12, 13, 14, 16)												8	4	6		3
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)														6	7	7
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)																3
Reading Comprehension for Information Retrieval (RCIR) (16)																3
Unbiased Learning to Ranking Evaluation Task (ULTRE) (16)																2

History of NTCIR-16 Tasks

Year	2013	2014	2016	2017	2019	2020	2022
Task/NTCIR round	10	11	12	13	14	15	16
Medical Natural Language Processing (MedNLP) (10, 11) -> MedNLPDoc (12) -> MedWeb (13) –> **Real-MedNLP**(16)	12	12	8	9			10
QA Lab for Entrance Exam (QALab) (11, 12, 13) -> QA Lab for Political Information (**QALab-PoliInfo**) (14, 15, 16)		11	12	11	13	14	12
Personal Lifelog Organisation & Retrieval (**Lifelog**) (12, 13, 14, 16)			8	4	6		3
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)					6	7	7
Dialogue Evaluation (**DialEval**) (15, 16)						7	4
Data Search (**Data Search**) (15, 16)						5	6
Session Search (**SS**) (16)							3
Reading Comprehension for Information Retrieval (**RCIR**) (16)							3
Unbiased Learning to Ranking Evaluation Task (**ULTRE**) (16)							2

Information Retrieval + NLP

Modern IR tasks from data to human

Data

Document

Human

Data Search 2	WWW-4	Lifelog-4
	SS	RCIR
	ULTRE	DialEval-2

Real-MedNLP	FinNum-3	QA Lab-PoliInfo-3

Medical

Finance

Politics

Deep NLP tasks in Specialized Domains

Online Proceedings are available

http://research.nii.ac.jp/ntcir/ntcir-16/

🏠 NTCIR Home

NTCIR 16
NTCIR 15
NTCIR 14
NTCIR 13

What's New

▸ archives

2022.06.14 🔲 new NTCIR-16 Online Proceedings are available here.

2022.05.19 NTCIR-16 Conference Online Registration opened!

NTCIR (NII Testbeds and Community for Information access Research) Project

NTCIR | CONTACT INFORMATION | NII

NTCIR

The 16th NTCIR Conference

Evaluation of Information Access Technologies

June 14-17, 2022
National Institute of Informatics, Tokyo, Japan

Top Page

Proceedings

NTCIR-16

Proceedings

NTCIR-16

Highlights of NTCIR-16 Conference

Schedule

- **June 14** JST(UTC+9)
  - Tutorial
- **June 15** JST(UTC+9)
  - Task Overviews
  - Keynotes
- **June 16** JST(UTC+9)
  - Task Sessions, Poster Sessions
- **June 17** JST(UTC+9)
  - Keynote, Invited Talks
  - Task Sessions, Poster Sessions

Tetsuya Sakai
(Waseda University, Japan)

You can watch the video at Lobby!

Keynotes

June 15

10:00 – 11:00 (JST, UTC+9)

Chengxiang Zhai
(University of Illinois at
Urbana-Champaign, USA)

June 15

20:00 – 21:00 (JST, UTC+9)

Ellen Voorhees
(NIST, USA)

June 17

17:00 – 18:00 (JST, UTC+9)

Falk Scholer
(RMIT University, Australia)

June 17   18:00 – (JST, UTC+9)

Nicola Ferro

Douglas Oard

Martha Larson

Gareth Jones

Poster Sessions

- **Poster 1 (QALab-PoliInfo-3, WWW-4, DialEval-2)**
  - 14:00 – 15:00, June 16 (JST, UTC+9)

- **Poster 2 (Lifelog-4, SS, RCIR)**
  - 18:30 – 19:30, June 16 (JST, UTC+9)
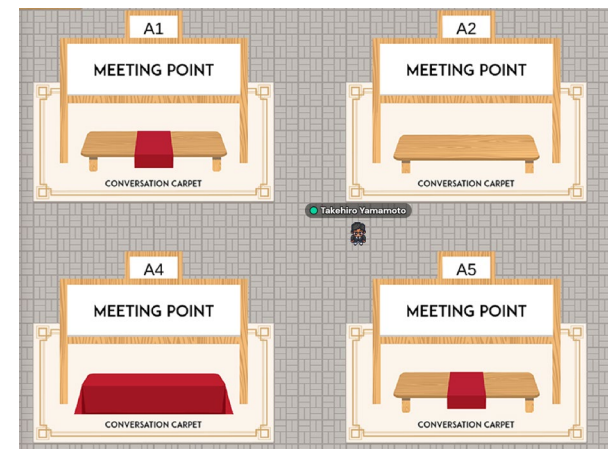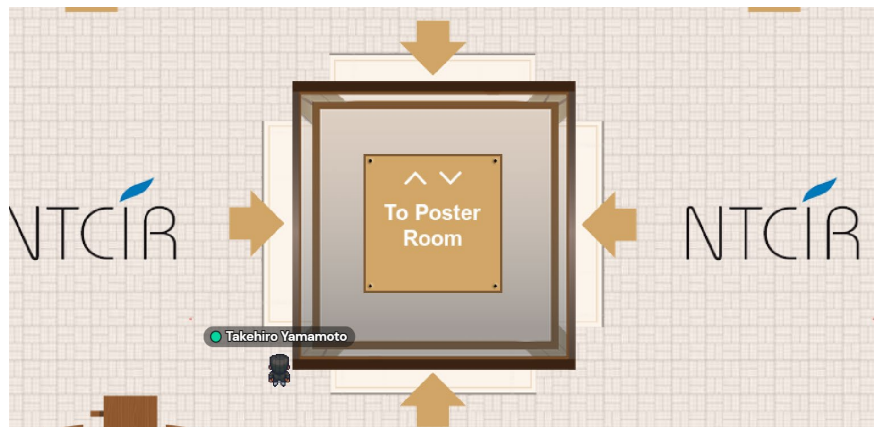
- **Poster 3 (FinNum-3, Data Search 2)**
  - 12:00 – 13:00, June 17 (JST, UTC+9)

- **Poster 4 (ULTRE, Real-MedNLP)**
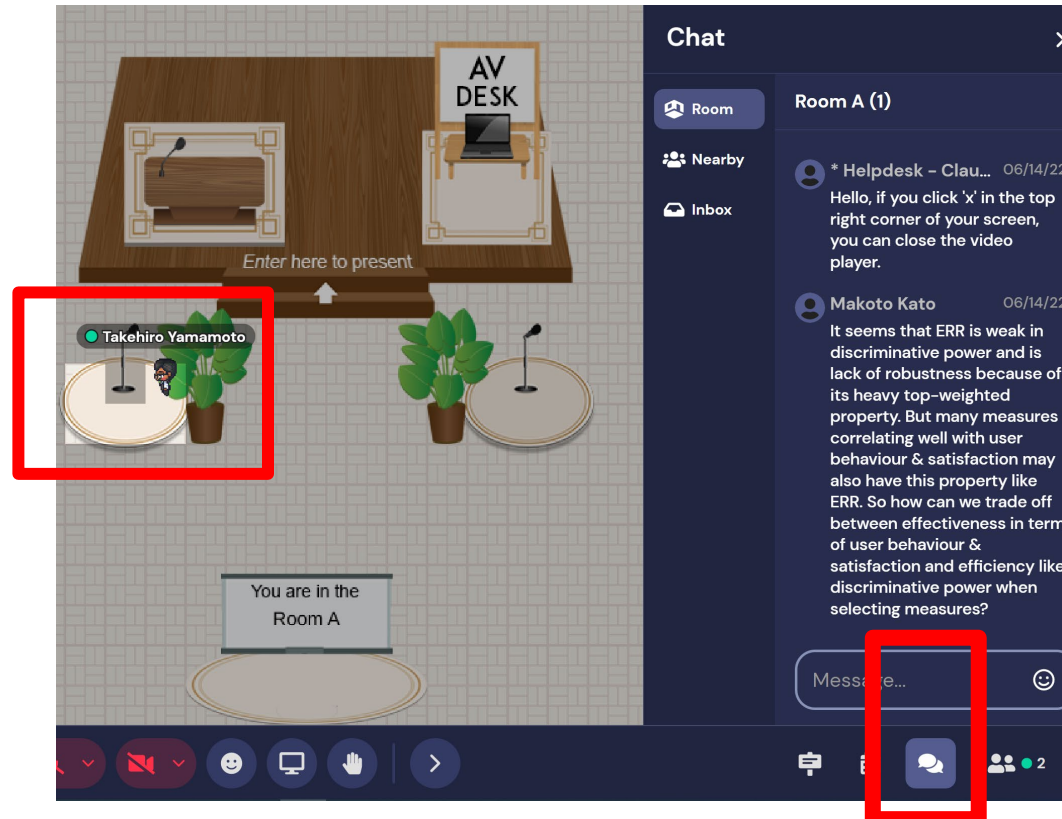 - 15:30 – 16:30, June 17 (JST, UTC+9)

Core time: 30 min.
- Odd number (A1, A3, ⋯)
  First 30 min.
- Even number (A2, A4, ⋯)
  Last 30 min.

Task Sessions

- **Overview from task organizers, participant presentations,**

- **Session Room A/B**

- Stand at Microphone to ask questions

- You can also use chat to ask questions

Press **X** to interact

POSTER DIRECTORY

Poster Session 1

[Poster Session 1] [Poster Session 2] [Poster Session 3] [Poster Session 4]

Location	Presenter	Title
A1	Fan Li, Tetsuya Sakai	[DialEval-2-1] RSLDE at the NTCIR-16 DialEval-2 Task
A2	Fei Ding, Kang Xin, Yunong Wu, Fuji Ren	[DialEval-2-2] TUA1 at the NTCIR-16 DialEval-2 Task
A3	Ting-Yun Hsiao, Yung-Wei Teng, Pei-Tz Chiu, Mike Tian-Jian Jiang, Min-Yuh Day	[DialEval-2-3] IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality and Nugget Detection
A4	Tao-Hsing Chang, Jian-He Chen	[DialEval-2-4] NKUST at the NTCIR-16 DialEval-2 Task
A5	Yuya Ubukata, Masaki Muraoka, Sijie Tao, Tetsuya Sakai	[WWW-4-1] SLWWW at the NTCIR-16 WWW-4 Task
A6	Kota Usuha, Kohei Shinden, Makoto P. Kato, Sumio Fujita	[WWW-4-2] KASYS at the NTCIR-16 WWW-4 Task
B1	Shenghao Yang, Haitao Li, Zhumin Chu, Jingtao Zhan, Yiqun Liu, Min Zhang, Shaoping Ma	[WWW-4-3] THUIR at the NTCIR-16 WWW-4 Task
B2	Ramon Ruiz Dolz	[QA Lab-PoliInfo-3-1] A Cascade Model for Argument Mining in Japanese Political Discussions: the QA Lab-PoliInfo-3 Case Study
B3	Yasuhiro Ogawa, Yugo Kato, Katsuhiko Toyama	[QA Lab-PoliInfo-3-10] nukl's QA System at the NTCIR-16 QA Lab-PoliInfo-3
B4	Kazuma Kadowaki, Shunsuke Onuma	[QA Lab-PoliInfo-3-11] JRIRD at the NTCIR-16 QA Lab-PoliInfo-3 Budget Argument Mining
B5	Daisa Nishihara, Hakuto Otatake, Kenji Yoshimura	[QA Lab-PoliInfo-3-2] fuys at the NTCIR-16 QA Lab-PoliInfo-3 Budget Argument Mining

PRESS 'X' FOR VIDEO

Day 3 Papers

[2] LIPI at the NTCIR-16 FinNum-3 Task: Ensembling transformer based models to detect in-claim numerals in Financial Conversations
Sohom Ghosh, Sudip Kumar Naskar

[7] UHGSIS at the NTCIR-16 Data Search 2 IR Subtask: A BERT-based Query Modification Approach
Moeri Okuda, Ryota Mibayashi, Takafumi Kawahara, Naoaki Matsumoto, Kenji Tanaka, Takehiro Yamamoto, Hiroaki Ohshima

[8] AMI team at the NTCIR-16 Real-MedNLP
Satoshi Hai, Shoji Nagayama, Atsushi Kojima

[9] SRCB at the NTCIR-16 Real-MedNLP Task
Yongwei Zhang, Rui Cheng, Lu Luo, Haifeng Gao, Shanshan Jiang, Bin Dong

[10] FRDC at the NTCIR-16 Real-MedNLP

UHGSIS at the NTCIR-16 Data Search 2 IR Subtask: A BERT-based Query Modification Approach

Moeri Okuda, Ryota Mibayashi, Takafumi Kawahara, Naoaki Matsumoto, Kenji Tanaka, Takehiro Yamamoto, Hiroaki Ohshima

- Announcements
- Can post/browse questions

- **Oral Presentation Award**
  - This award is given to **oral presentations** that are excellent in novelty, reproducibility, validity, presentation quality, and quality of Q&A
  - The award winner is determined based on recommendation by the **task organizers**
  - **Thank you to the task organizers for their cooperation**

- **Poster Presentation Awards**
  - This award is given to excellent **poster presentations**.
  - The award winner is determined based on a **vote of all participants**.
  - Please participate in the voting for **each poster session**.

- **The winners will be announced on the closing session on Day 4**

Quick Tips in Gather.town

Press "X"
- Open/close video,
- Join zoom,
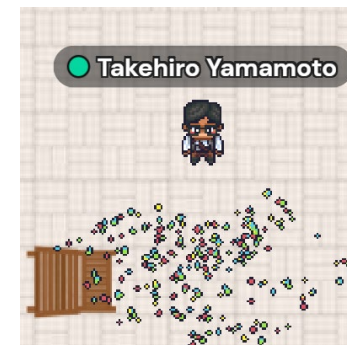- Open schedule…
- etc.

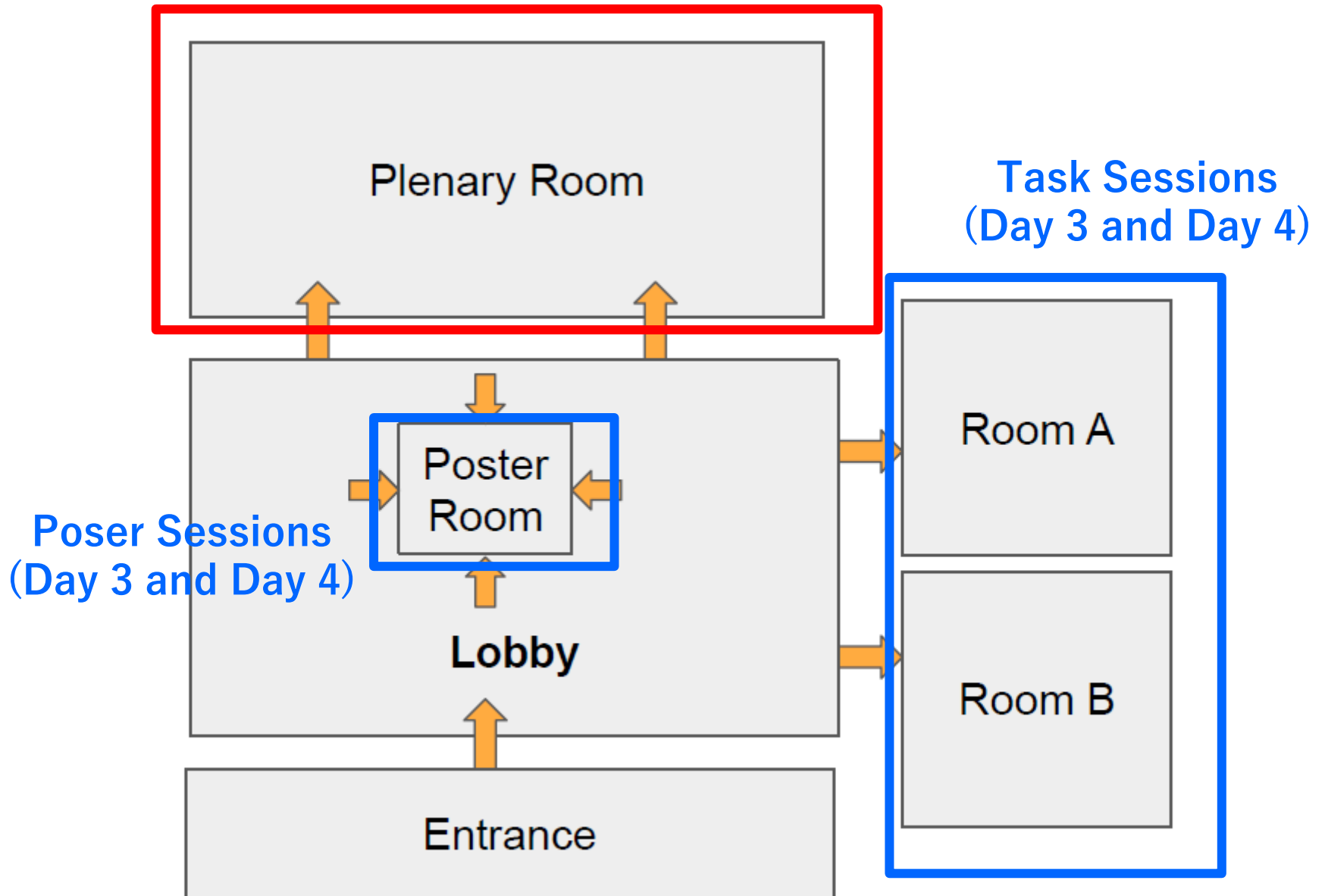**Press "G"** to go through crowds

Other tips…

Press "Z" for dancing

Press "F" for Confetti

Opening, Keynotes, and Task Overview Presentations

Plenary Room

Task Sessions (Day 3 and Day 4)

Room A

Room B

Poser Sessions (Day 3 and Day 4)

Poster Room

Lobby

Entrance

Summary

Summary

- **16th round of NTCIR**

- **10 tasks: 6 core and 4 pilot tasks**

  – Cover a wide range of IR and NLP tasks!

- **Tutorial, Keynotes, Task Overview Presentations, Participants' Poster and Oral Presentations**

- **Online Proceedings are available on the Website**

- **Participants' Presentations and Posters are also available in the gather.town platform**

Virtual

Enjoy NTCIR-16 Conference!