WAT 2024

The 11th Workshop on Asian Translation

Proceedings of the Workshop

November 16, 2024

The WAT organizers gratefully acknowledge the support from the following sponsors.

Order copies of this and other ACL proceedings from:

Preface

Many Asian countries are rapidly growing these days and the importance of communicating and exchanging the information with these countries has intensified. To satisfy the demand for communication among these countries, machine translation technology is essential.

Machine translation technology has rapidly evolved recently and it is seeing practical use especially between European languages. However, the translation quality of Asian languages is not that high compared to that of European languages, and machine translation technology for these languages has not reached a stage of proliferation yet. This is not only due to the lack of the language resources for Asian languages but also due to the lack of techniques to correctly transfer the meaning of sentences from/to Asian languages. Consequently, a place for gathering and sharing the resources and knowledge about Asian language translation is necessary to enhance machine translation research for Asian languages.

The Conference on Machine Translation (WMT), the world's largest machine translation conference, mainly targets on European language. The International Workshop on Spoken Language Translation (IWSLT) has spoken language translation tasks for some Asian languages using TED talk data, but there is no task for written language. The Workshop on Asian Translation (WAT) is an open machine translation evaluation campaign focusing on Asian languages. WAT gathers and shares the resources and knowledge of Asian language translation to understand the problems to be solved for the practical use of machine translation technologies among all Asian countries. WAT is unique in that it is an open innovation platform": the test data is fixed and open, so participants can repeat evaluations on the same data and confirm changes in translation accuracy over time. WAT has no deadline for the automatic translation quality evaluation (continuous evaluation), so participants can submit translation results at any time.

Following the success of the previous WAT workshops (WAT2014 - WAT2023), WAT2024 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 11th WAT, we have MULTIINDIC22MT TASK, ENGLISH-TO-LOWRES MULTIMODAL MT TASK, NON-REPETITIVE TASK, PATENT TASK. This year, the shared tasks were conducted under WMT2024, which was held in the same venue as WAT2024, for the participants' convenience. The system description papers are archived in the WMT2024 proceedings.

In addition to the shared tasks, WAT2024 also features research papers on topics related to machine translation, especially for Asian languages. We received 11 research papers submitted, and the program committee accepted 6 research papers.

We would like to thank all the authors who submitted papers. We express our deepest gratitude to the committee members for their timely reviews. We also thank the EMNLP 2024 organizers for their help with administrative matters.

WAT 2024 Organizers

Organizing Committee

Organizers

Toshiaki Nakazawa, The University of Tokyo, Japan
Isao Goto, Ehime University, Japan
Hidaya Mino, Japan Broadcasting Corporation (NHK), Japan
Kazutaka Kinugawa, Japan Broadcasting Corporation (NHK), Japan
Chenhui Chu, Kyoto University, Japan
Haiyue Song, National Institute of Information and Communications Technology (NICT), Japan
Raj Dabre, National Institute of Information and Communications Technology (NICT), Japan
Shohei Higashiyama, National Institute of Information and Communications Technology (NICT), Japan
Anoop Kunchookuttan, Microsoft AI and Research, India
Shantipriya Parida, Silo AI, Finland
Ondřej Bojar, Charles University, Prague, Czech Republic
Sadao Kurohashi, National Institute of Informatics, Japan
Pushpak Bhattacharyya, Indian Institute of Technology Patna (IITP), India

Technical Collaborators

Luis Fernando D'Haro, Universidad Politécnica de Madrid, Spain
Rafael E. Banchs, Nanyang Technological University, Singapore
Haizhou Li, National University of Singapore, Singapore
Chen Zhang, National University of Singapore, Singapore

Program Committee

Program Committee

    Raj Dabre, NICT
    Shohei Higashiyama, NICT
    Kenji Imamura, NICT
    Chao-Hong Liu, Potamu Research Limited
    Hideya Mino, NHK
    Takashi Ninomiya, Ehime University
    Shantipriya Parida, Silo AI
    Katsuhito Sudoh, Nara Women's University
    Masao Utiyama, NICT
    Isao Goto, Ehime University

Panelists

    Min Zhang, Soochow University, China
    Thepchai Supnithi, National Electronics and Computer Technology Center (NECTEC), Thailand
    Kozo Moriguchi, Kawamura International Co., Ltd., Japan

Table of Contents

Program

Saturday, November 16, 2024

09:00 - 09:05 *Welcome*

09:05 - 10:00 *Panel Discussion*: *"Machine Translation of Asian Languages in the LLM Era"*

10:00 - 10:40 *Research Paper I*

       Machine Translation Of Marathi Dialects: A Case Study Of Kadodi
Raj Dabre, Mary Dabre and Teresa Pereira

       AI-Tutor: Interactive Learning of Ancient Knowledge from Low-Resource Languages
Siddhartha Dalal, Rahul Aditya, Vethavikashini Chithrra Raghuram and Prahlad Koratamaddi

10:40 - 11:10 *Break*

11:10 - 12:10 *Research Paper II*

       An Empirical Study of Multilingual Vocabulary for Neural Machine Translation Models
Kenji Imamura and Masao Utiyama

       Are Large Language Models State-of-the-art Quality Estimators for Machine Translation of User-generated Content?
Shenbin Qian, Constantin Orasan, Diptesh Kanojia and Félix Do Carmo

       Creative and Context-Aware Translation of East Asian Idioms with GPT-4
Kenan Tang, Peiyang Song, Yao Qin and Xifeng Yan

12:10 - 12:15 *Closing*