

# **Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024)**

Languages of Latin America

Demonstrations Session

Mexico City, Mexico  
16-21 June 2024

ISBN: 978-1-7138-9896-2

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2024) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>TOPICAL: TOPIC Pages Automagically</i>	
John Michael Giorgi, Amanpreet Singh, Doug Downey, Sergey Feldman and Lucy Lu Wang . . .	1
<i>Low-code LLM: Graphical User Interface over Large Language Models</i>	
Yuzhe Cai, Shaoguang Mao, Wenshan Wu, Zehua Wang, Yaobo Liang, Tao Ge, Chenfei Wu, WangYou WangYou, Ting Song, Yan Xia, Nan Duan and Furu Wei . . . . .	12
<i>EdTec-QBuilder: A Semantic Retrieval Tool for Assembling Vocational Training Exams in German Language</i>	
Alonso Palomino, Andreas Fischer, Jakub Kuzilek, Jarek Nitsch, Niels Pinkwart and Benjamin Paassen . . . . .	26
<i>DIALIGHT: Lightweight Multilingual Development and Evaluation of Task-Oriented Dialogue Systems with Large Language Models</i>	
Songbo Hu, Xiaobin Wang, Moy Yuan, Anna Korhonen and Ivan Vulić . . . . .	36
<i>RTSUM: Relation Triple-based Interpretable Summarization with Multi-level Saliency Visualization</i>	
Seonglae Cho, Myungha Jang, Jinyoung Yeo and Dongha Lee . . . . .	53
<i>Edu-ConvoKit: An Open-Source Library for Education Conversation Data</i>	
Rose E Wang and Dorottya Demszky . . . . .	61
<i>jp-evalb: Robust Alignment-based PARSEVAL Measures</i>	
Jungyeul Park, Junrui Wang, Eunkyul Leah Jo and Angela Yoonseo Park . . . . .	70
<i>OpinionGPT: Modelling Explicit Biases in Instruction-Tuned LLMs</i>	
Patrick Haller, Ansar Aynetdinov and Alan Akbik . . . . .	78
<i>ATLAS: A System for PDF-centric Human Interaction Data Collection</i>	
Alexa Siu, Zichao Wang, Joshua Hoeflich, Naman Kapasi, Ani Nenkova and Tong Sun . . . . .	87
<i>BeLeaf: Belief Prediction as Tree Generation</i>	
John Murzaku and Owen Rambow . . . . .	97
<i>QueryExplorer: An Interactive Query Generation Assistant for Search and Exploration</i>	
Kaustubh Dhole, Shivam Bajaj, Ramraj Chandradevan and Eugene Agichtein . . . . .	107
<i>LMFlow: An Extensible Toolkit for Finetuning and Inference of Large Foundation Models</i>	
Shizhe Diao, Rui Pan, Hanze Dong, KaShun Shum, Jipeng Zhang, Wei Xiong and Tong Zhang	116
<i>DOCMaster: A Unified Platform for Annotation, Training, &amp; Inference in Document Question-Answering</i>	
Alex Nguyen, Zilong Wang, Jingbo Shang and Dheeraj Mekala . . . . .	128
<i>RedCoast: A Lightweight Tool to Automate Distributed Training of LLMs on Any GPU/TPUs</i>	
Bowen Tan, Yun Zhu, Lijuan Liu, Hongyi Wang, Yonghao Zhuang, Jindong Chen, Eric P. Xing and Zhiting Hu . . . . .	137
<i>Concept Over Time Analysis: Unveiling Temporal Patterns for Qualitative Data Analysis</i>	
Tim Fischer, Florian Schneider, Robert Geislinger, Florian Helfer, Gertraud Koch and Chris Biemann . . . . .	148

<i>pyvene: A Library for Understanding and Improving PyTorch Models via Interventions</i> Zhengxuan Wu, Atticus Geiger, Aryaman Arora, Jing Huang, Zheng Wang, Noah Goodman, Christopher D Manning and Christopher Potts .....	158
<i>Newspaper Signaling for Crisis Prediction</i> Prajvi Saxena, Sabine Janzen and Wolfgang Maass .....	166
<i>FastFit: Fast and Effective Few-Shot Text Classification with a Multitude of Classes</i> Asaf Yehudai and Elron Bandel .....	174
<i>AgentQuest: A Modular Benchmark Framework to Measure Progress and Improve LLM Agents</i> Luca Gioacchini, Giuseppe Siracusano, Davide Sanvito, Kiril Gashteovski, David Friede, Roberto Bifulco and Carolin Lawrence .....	185
<i>ZhuJiu-Knowledge: A Fairer Platform for Evaluating Multiple Knowledge Types in Large Language Models</i> Pengfan DU, Sirui Liang, Baoli Zhang, Pengfei Cao, Yubo Chen, Kang Liu and Jun Zhao ...	194
<i>Unitxt: Flexible, Shareable and Reusable Data Preparation and Evaluation for Generative AI</i> Elron Bandel, Yotam Perlitz, Elad Venezian, Roni Friedman, Ofir Arviv, Matan Orbach, Shachar Don-Yehiya, Dafna Sheinwald, Ariel Gera, Leshem Choshen, Michal Shmueli-Scheuer and Yoav Katz	207