

2nd Workshop on Natural Language Generation, Evaluation and Metrics (GEM 2022)

Abu Dhabi, United Arab Emirates and Online
7 December 2022

ISBN: 978-1-7138-9111-6

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© (2022) 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

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
Web: www.proceedings.com

Table of Contents

<i>Improving abstractive summarization with energy-based re-ranking</i> Diogo Pernes, Afonso Mendes and André F. T. Martins	1
<i>Task-driven augmented data evaluation</i> Olga Golovneva, Pan Wei, Khadige Abboud, Charith Peris, Lizhen Tan and Haiyang Yu	18
<i>Generating Coherent Narratives with Subtopic Planning to Answer How-to Questions</i> Pengshan Cai, Mo Yu, Fei Liu and Hong Yu	26
<i>Weakly Supervised Context-based Interview Question Generation</i> Samiran Pal, Kaamraan Khan, Avinash Kumar Singh, Subhasish Ghosh, Tapas Nayak, Girish Palshikar and Indrajit Bhattacharya	43
<i>Analyzing Multi-Task Learning for Abstractive Text Summarization</i> Frederic Thomas Kirstein, Jan Philip Wahle, Terry Ruas and Bela Gipp	54
<i>CLSE: Corpus of Linguistically Significant Entities</i> Aleksandr Chuklin, Justin Zhao and Mihir Kale	78
<i>Revisiting text decomposition methods for NLI-based factuality scoring of summaries</i> John Glover, Federico Fancellu, Vasudevan Jagannathan, Matthew R. Gormley and Thomas Schaaf	97
<i>Semantic Similarity as a Window into Vector- and Graph-Based Metrics</i> Wai Ching Leung, Shira Wein and Nathan Schneider	106
<i>Towards In-Context Non-Expert Evaluation of Reflection Generation for Counselling Conversations</i> Zixiu Wu, Simone Balloccu, Rim Helaoui, Diego Reforgiato Recupero and Daniele Riboni	116
<i>WikiOmnia: filtration and evaluation of the generated QA corpus on the whole Russian Wikipedia</i> Dina Pisarevskaya and Tatiana Shavrina	125
<i>Evaluation of Response Generation Models: Shouldn't It Be Shareable and Replicable?</i> Seyed Mahed Mousavi, Gabriel Roccabruna, Michela Lorandi, Simone Caldarella and Giuseppe Riccardi	136
<i>Enhancing and Evaluating the Grammatical Framework Approach to Logic-to-Text Generation</i> Eduardo Calò, Elze van der Werf, Albert Gatt and Kees van Deemter	148
<i>Controllable Text Generation for All Ages: Evaluating a Plug-and-Play Approach to Age-Adapted Dialogue</i> Lennert Jansen, Štěpán Lars Laichter, Arabella Sinclair, Margot van der Goot, Raquel Fernandez and Sandro Pezzelle	172
<i>Template-based Contact Email Generation for Job Recommendation</i> Qiuchi Li and Christina Lioma	189
<i>Are Abstractive Summarization Models truly 'Abstractive'? An Empirical Study to Compare the two Forms of Summarization</i> Vinayshekhhar Bannihatti Kumar and Rashmi Gangadharaiyah	198
<i>Transfer learning for multilingual vacancy text generation</i> Anna Lorincz, David Graus, Dor Lavi and Joao Lebre Magalhaes Pereira	207

<i>Plug-and-Play Recipe Generation with Content Planning</i> Yinhong Liu, Yixuan Su, Ehsan Shareghi and Nigel Collier	223
<i>Towards Attribute-Entangled Controllable Text Generation: A Pilot Study of Blessing Generation</i> Shulin Huang, Shirong Ma, Yinghui Li, Li Yangning, Shiyang Lin, Haitao Zheng and Ying Shen	235
<i>Unsupervised Token-level Hallucination Detection from Summary Generation By-products</i> Andreas Marfurt and James Henderson	248
<i>A Corpus and Evaluation for Predicting Semi-Structured Human Annotations</i> Andreas Marfurt, Ashley Thornton, David Sylvan, Lonneke van der Plas and James Henderson	262
<i>T5QL: Taming language models for SQL generation</i> Samuel David Arcadinho, David Aparicio, Hugo Veiga and Antonio Alegria	276
<i>Human perceiving behavior modeling in evaluation of code generation models</i> Sergey V. Kovalchuk, Vadim Lomshakov and Artem Aliev	287
<i>Nearest Neighbor Language Models for Stylistic Controllable Generation</i> Severino Trotta, Lucie Flek and Charles Welch	295
<i>On reporting scores and agreement for error annotation tasks</i> Maja Popović and Anya Belz	306
<i>Answerability: A custom metric for evaluating chatbot performance</i> Pranav Gupta, Anand A. Rajasekar, Amisha Patel, Mandar Kulkarni, Alexander Sunell, Kyung Kim, Krishnan Ganapathy and Anusua Trivedi	316
<i>Improved Evaluation of Automatic Source Code Summarisation</i> Jesse Phillips, David Bowes, Mahmoud El-Haj and Tracy Hall	326
<i>Most NLG is Low-Resource: here's what we can do about it</i> David M. Howcroft and Dimitra Gkatzia	336
<i>GiCCS: A German in-Context Conversational Similarity Benchmark</i> Shima Asaadi, Zahra Kolagar, Alina Liebel and Alessandra Zarcone	351
<i>Control Prefixes for Parameter-Efficient Text Generation</i> Jordan Clive, Kris Cao and Marek Rei	363
<i>A Survey of Recent Error Annotation Schemes for Automatically Generated Text</i> Rudali Huidrom and Anya Belz	383
<i>What's in a (dataset's) name? The case of BigPatent</i> Silvia Casola, Alberto Lavelli and Horacio Saggion	399
<i>Measuring the Measuring Tools: An Automatic Evaluation of Semantic Metrics for Text Corpora</i> George Kour, Samuel Ackerman, Eitan Daniel Farchi, Orna Raz, Boaz Carmeli and Ateret Anaby Tavor	405
<i>Multilingual Social Media Text Generation and Evaluation with Few-Shot Prompting</i> Mack Blackburn	417
<i>Assessing Inter-metric Correlation for Multi-document Summarization Evaluation</i> Michael Ridenour, Ameeta Agrawal and Olubusayo Olabisi	428

<i>Factual Error Correction for Abstractive Summaries Using Entity Retrieval</i> Hwanhee Lee, Cheoneum Park, Seunghyun Yoon, Trung Bui, Franck Dernoncourt, Juae Kim and Kyomin Jung	439
<i>Coherent Long Text Generation by Contrastive Soft Prompt</i> Guandan Chen, Jiashu Pu, Yadong Xi and Rongsheng Zhang	445
<i>Error Analysis of ToTTo Table-to-Text Neural NLG Models</i> Barkavi Sundararajan, Somayajulu Sripada and Ehud Reiter	456
<i>Improving Dialogue Act Recognition with Augmented Data</i> Khyati Mahajan, Soham Parikh, Quaizar Vohra, Mitul Tiwari and Samira Shaikh	471
<i>Do Decoding Algorithms Capture Discourse Structure in Multi-Modal Tasks? A Case Study of Image Paragraph Generation</i> Nikolai Ilinykh and Simon Dobnik	480
<i>20Q: Overlap-Free World Knowledge Benchmark for Language Models</i> Maxime De Bruyn, Ehsan Lotfi, Jeska Buhmann and Walter Daelemans	494
<i>What Was Your Name Again? Interrogating Generative Conversational Models For Factual Consistency Evaluation</i> Ehsan Lotfi, Maxime De Bruyn, Jeska Buhmann and Walter Daelemans	509
<i>Narrative Why-Question Answering: A Review of Challenges and Datasets</i> Emil Kalbaliyev and Kairit Sirts	520
<i>Exploring a POS-based Two-stage Approach for Improving Low-Resource AMR-to-Text Generation</i> Marco Antonio Sobrevilla Cabezudo and Thiago Pardo	531
<i>What Makes Data-to-Text Generation Hard for Pretrained Language Models?</i> Moniba Keymanesh, Adrian Benton and Mark Dredze	539
<i>Don't Say What You Don't Know: Improving the Consistency of Abstractive Summarization by Constraining Beam Search</i> Daniel King, Zejiang Shen, Nishant Subramani, Daniel S. Weld, Iz Beltagy and Doug Downey	555