EMNLP 2021 Workshop

Proceedings of the 3rd Workshop on Machine Reading for Question Answering

November 10th, 2021

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-95-4

Message from the Organizers

Our workshop brings together researchers studying machine reading for question answering (MRQA). MRQA has emerged as an important testbed for evaluating how computer systems understand natural language, as well as a crucial technology for applications such as search engines and dialog systems. In recent years, MRQA systems have become much more accurate, and are even capable of retrieving evidence documents on the fly or answering without retrieved documents. Datasets and models have been developed to target many different aspects of the problem, including multi-hop reasoning, numerical reasoning, or commonsense reasoning.

Despite this progress, there are still many important desiderata that most MRQA systems do not adequately consider: multilinguality and interpretability. In the 3rd MRQA workshop, we therefore focus on these two emerging and crucial aspects of question answering models.

Systems today are predominantly evaluated by measuring accuracy on English benchmarks, yet an ideal question answering system would support a diverse range of languages. With recent developments of multilingual question answering datasets, it is timely to study how MRQA models can be designed to support typologically diverse languages.

Many systems produce correct answers for the wrong reason and are unable to explain their predictions. Given the opaque nature of modern large-scale pre-trained neural models, it is important to study how MRQA systems can offer users a way to trust (or not trust) an otherwise black-box model's predictions, as well as offer practitioners ways to diagnose critical modeling issues or dataset biases.

As in past years, we sought paper submissions of previously unpublished work. To reflect our focus on our two themes, we had separate tracks for multilinguality and interpretability-related papers, as well as a general research track. Across these three tracks, we received 21 total paper submissions after withdrawals – 14 for the general research track, 5 for the multilingual track, and 2 for the interpretability track. While the submission counts have decreased from last year, we found the average quality of submitted papers to be higher than previous years. After discussion among the organizers, we have accepted a total of 16 papers and awarded one best paper and two honorable mention papers. We also have accepted 23 non-archival submissions that were accepted at other related conferences (such as papers accepted at the main conference/findings of ACL, EMNLP, SIGIR) to be presented at our workshop. Our final program therefore includes 39 papers, of which 16 papers are included in these proceedings.

We are excited to host six stellar invited speakers. In the morning session, Reut Tsarfaty, Jon Clark, and Yiming Cui will give talks on multilinguality in question answering; in the afternoon session, Jonathan Berant, Marco Tulio Ribeiro, and Hannaneh Hajishirzi will give talks on interpretability in question answering. We thank these speakers, our program committee, the EMNLP workshop chairs, and our sponsors, Baidu and Facebook, for helping to make this workshop possible.

Organizing Committee

Adam Fisch, MIT
Alon Talmor, Tel Aviv University
Danqi Chen, Princeton University
Eunsol Choi, The University of Texas at Austin
Minjoon Seo, Naver & KAIST
Patrick Lewis, Facebook & University College London
Robin Jia, University of Southern California
Sewon Min, University of Washington

Program Committee

Akari Asai
Danish Contractor
Douwe Kiela
Eric Wallace
Gautier Izacard
Huan Sun
Jifan Chen
Jing Liu
Jinhyuk Lee
Jonathan Clark
Jonathan Herzig
Kai Sun
Kenton Lee
Kevin Lin
Matt Gardner
Matthew Lamm
Max Bartolo
Mor Geva
Nan Duan
Nitish Gupta
Panupong Pasupat
Pedro Rodriguez
Peng Qi
Rajarshi Das
Rodrigo Nogueira
Shayne Longpre
Shuohang Wang
Thomas Kwiatkowski
Tong Wang
Tushar Khot
Xiaodong Liu
Xinya Du
Yichen Jiang
Yizhong Wang
Yuxiang Wu

Invited Speaker

Reut Tsarfaty, Bar Illan University
Jon Clark, Google
Yiming Cui, Joint Laboratory of HIT and iFLYTEK Research (HFL)
Jonathan Berant, Tel Aviv University, Allen Institute for AI
Marco Tulio Ribeiro, Microsoft Research
Hannah Hajishirzi, University of Washington, Allen Institute for Artificial Intelligence

Table of Contents

Conference Program

Wednesday, November 10, 2021

9:00–9:15 ***Opening Remarks***

9:15–11:30 **Multilingual QA Invited Talk Session**

9:15–9:45 ***Invited Talk 1 - Reut Tsarfaty***

9:45–10:15 ***Invited Talk 2 - Jon Clark***

10:15–10:45 ***Invited Talk 3 - Yiming Cui***

10:45–11:30 ***Panel Discussion on Multilingual QA***

11:30–12:30 ***Lunch break***

12:30–13:10 **Best Paper Talk Session**

12:30–12:44 *MFAQ: a Multilingual FAQ Dataset*
Maxime De Bruyn, Ehsan Lotfi, Jeska Buhmann and Walter Daelemans

12:44–12:57 *Rethinking the Objectives of Extractive Question Answering*
Martin Fajcik, Josef Jon and Pavel Smrz

12:57–13:10 *What Would it Take to get Biomedical QA Systems into Practice?*
Gregory Kell, Iain Marshall, Byron Wallace and Andre Jaun

13:10–14:10 Poster Session (archival track)

13:10–14:10 *GermanQuAD and GermanDPR: Improving Non-English Question Answering and Passage Retrieval*
Timo Möller, Julian Risch and Malte Pietsch

13:10–14:10 *Zero-Shot Clinical Questionnaire Filling From Human-Machine Interactions*
Farnaz Ghassemi Toudeshki, Philippe Jolivet, Alexandre Durand-Salmon and Anna Liednikova

13:10–14:10 *Can Question Generation Debias Question Answering Models? A Case Study on Question–Context Lexical Overlap*
Kazutoshi Shinoda, Saku Sugawara and Akiko Aizawa

13:10–14:10 *What Can a Generative Language Model Answer About a Passage?*
Douglas Summers-Stay, Claire Bonial and Clare Voss

13:10–14:10 *Multi-modal Retrieval of Tables and Texts Using Tri-encoder Models*
Bogdan Kostić, Julian Risch and Timo Möller

13:10–14:10 *Eliciting Bias in Question Answering Models through Ambiguity*
Andrew Mao, Naveen Raman, Matthew Shu, Eric Li, Franklin Yang and Jordan Boyd-Graber

13:10–14:10 *Bilingual Alignment Pre-Training for Zero-Shot Cross-Lingual Transfer*
Ziqing Yang, Wentao Ma, Yiming Cui, Jiani Ye, Wanxiang Che and Shijin Wang

13:10–14:10 *ParaShoot: A Hebrew Question Answering Dataset*
Omri Keren and Omer Levy

13:10–14:10 *Unsupervised Multiple Choices Question Answering: Start Learning from Basic Knowledge*
Chi-Liang Liu and Hung-yi Lee

13:10–14:10 *GANDALF: a General Character Name Description Dataset for Long Fiction*
Fredrik Carlsson, Magnus Sahlgren, Fredrik Olsson and Amaru Cuba Gyllensten

13:10–14:10 *Investigating Post-pretraining Representation Alignment for Cross-Lingual Question Answering*
Fahim Faisal and Antonios Anastasopoulos

Wednesday, November 10, 2021 (continued)

13:10–14:10 *Semantic Answer Similarity for Evaluating Question Answering Models*
Julian Risch, Timo Möller, Julian Gutsch and Malte Pietsch

13:10–14:10 *Simple and Efficient ways to Improve REALM*
Vidhisha Balachandran, Ashish Vaswani, Yulia Tsvetkov and Niki Parmar

13:10–14:10 Poster Session (non-archival track)

13:10–14:10 *Synthetic Target Domain Supervision for Open Retrieval QA*
Revanth Gangi Reddy, Bhavani Iyer, Md Arafat Sultan, Rong Zhang, Avirup Sil, Vittorio Castelli, Radu Florian, Salim Roukos

13:10–14:10 *Entity-based Knowledge Conflicts in Question Answering*
Shayne Longpre, Kartik Perisetla, Anthony Chen, Nikhil Ramesh, Chris Dubois, Sameer Singh

13:10–14:10 *Mitigating False-Negative Contexts in Multi-Document Question Answering with Retrieval Marginalization*
Ansong Ni, Matt Gardner, Pradeep Dasigi

13:10–14:10 *Generative Context Pair Selection for Multi-hop Question Answering*
Dheeru Dua,Cicero Nogueira dos Santos,Patrick Ng,Ben Athiwaratkun,Bing Xiang,Matt Gardner,Sameer Singh

13:10–14:10 *Learning with Instance Bundles for Reading Comprehension*
Dheeru Dua, Pradeep Dasigi,Sameer Singh,Matt Gardner

13:10–14:10 *Can NLI Models Verify QA Systems' Predictions?*
Jifan Chen, Eunsol Choi, Greg Durrett

13:10–14:10 *Knowing More About Questions Can Help: Improving Calibration in Question Answering*
Shujian Zhang, Chengyue Gong and Eunsol Choi

13:10–14:10 *RocketQA: An Optimized Training Approach to Dense Passage Retrieval for Open-Domain Question Answering*
Yingqi Qu, Yuchen Ding, Jing Liu, Kai Liu, Ruiyang Ren, Wayne Xin Zhao, Daxiang Dong, Hua Wu and Haifeng Wang

13:10–14:10 *Weakly Supervised Pre-Training for Multi-Hop Retriever*
Yeon Seonwoo, Sang-Woo Lee, Ji-Hoon Kim, Jung-Woo Ha and Alice Oh

13:10–14:10 *ReasonBert: Pre-trained to Reason with Distant Supervision*
Xiang Deng, Yu Su, Alyssa Lees, You Wu, Cong Yu and Huan Sun

Wednesday, November 10, 2021 (continued)

13:10–14:10 *Question Answering over Electronic Devices: A New Benchmark Dataset and a Multi-Task Learning based QA Framework*
Abhilash Nandy, Soumya Sharma, Shubham Maddhashiya, Kapil Sachdeva, Pawan Goyal and NIloy Ganguly

13:10–14:10 *Do We Know What We Don't Know? Studying Unanswerable Questions beyond SQuAD 2.0*
Elior Sulem, Jamaal Hay and Dan Roth

13:10–14:10 *Relation-Guided Pre-Training for Open-Domain Question Answering*
Ziniu Hu, Yizhou Sun and Kai-Wei Chang

13:10–14:10 *Beyond Reptile: Meta-Learned Dot-Product Maximization between Gradients for Improved Single-Task Regularization*
Akhil Kedia, Sai Chetan Chinthakindi and Wonho Ryu

13:10–14:10 *SD-QA: Spoken Dialectal Question Answering for the Real World*
Fahim Faisal, Sharlina Keshava, Md Mahfuz Ibn Alam and Antonios Anastasopoulos

13:10–14:10 *When Retriever-Reader Meets Scenario-Based Multiple-Choice Questions*
ZiXian Huang, Ao Wu, Yulin Shen, Gong Cheng and Yuzhong Qu

13:10–14:10 *Winnowing Knowledge for Multi-choice Question Answering*
Yeqiu Li, Bowei Zou, Zhifeng Li, Ai Ti Aw, Yu Hong and Qiaoming Zhu

13:10–14:10 *Extract, Integrate, Compete: Towards Verification Style Reading Comprehension*
Chen Zhang, Yuxuan Lai, Yansong Feng and Dongyan Zhao

13:10–14:10 *Reference-based Weak Supervision for Answer Sentence Selection using Web Data*
Vivek Krishnamurthy, Thuy Vu and Alessandro Moschitti

13:10–14:10 *NOAHQA: Numerical Reasoning with Interpretable Graph Question Answering Dataset*
Qiyuan Zhang, Lei Wang, SICHENG YU, Shuohang Wang, Yang Wang, Jing Jiang and Ee-Peng Lim

13:10–14:10 *Improving Numerical Reasoning Skills in the Modular Approach for Complex Question Answering on Text*
Xiao-Yu Guo, Yuan-Fang Li and Gholamreza Haffari

13:10–14:10 *R2-D2: A Modular Baseline for Open-Domain Question Answering*
Martin Fajcik, Martin Docekal, Karel Ondrej and Pavel Smrz

13:10–14:10 *AutoEQA: Auto-Encoding Questions for Extractive Question Answering*
Stalin Varanasi, Saadullah Amin and Guenter Neumann

14:10–14:30 ***Break***

Wednesday, November 10, 2021 (continued)

14:30–16:45 Interpretability in QA Invited Talk Session

14:30–15:00 ***Invited Talk 4 - Jonathan Berant***

15:00–15:30 ***Invited Talk 5 - Marco Tulio Ribeiro***

15:30–16:00 ***Invited Talk 6 - Hannaneh Hajishirzi***

16:00–16:45 ***Panel Discussion on Interpretability in QA***

16:45–17:00 ***Closing Remarks***