

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their feedback. This research was partially supported by the Netherlands Organisation for Scientific Research (NWO) under project nr CI-14-25, the NWO Innovative Research Incentives Scheme Vidi (016.Vidi.189.039), the NWO Smart Culture - Big Data / Digital Humanities (314-99-301), the H2020-EU.3.4. - SOCIETAL CHALLENGES - Smart, Green And Integrated Transport (814961), the Google Faculty Research Awards program, and the Innovation Center for AI (ICAI). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

CODE AND DATA

To facilitate reproducibility, we share the resources used in this paper at <https://github.com/nickvosk/sigir2020-query-resolution>.

REFERENCES

- [1] Nasreen Abdul-jaleel, James Allan, W Bruce Croft, O Diaz, Leah Larkey, Xiaoyan Li, Mark D Smucker, and Courtney Wade. 2004. UMass at TREC 2004: Novelty and HARD. In *TREC*. NIST.
- [2] Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. 2018. MS MARCO: A Human Generated MACHine Reading COmprehension Dataset. *arXiv preprint arXiv:1611.09268* (2018).
- [3] Nicholas J Belkin. 1980. Anomalous States of Knowledge as A Basis for Information Retrieval. *Canadian Journal of Information Science* 5, 1 (1980), 133–143.
- [4] Nicholas J Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. 1995. Cases, Scripts, and Information-seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications* 9, 3 (1995), 379–395.
- [5] Ben Carterette, Paul Clough, Mark Hall, Evangelos Kanoulas, and Mark Sanderson. 2016. Evaluating Retrieval over Sessions: The TREC Session Track 2011-2014. In *SIGIR*. ACM, 685–688.
- [6] Ben Carterette, Evangelos Kanoulas, Mark Hall, and Paul Clough. 2014. *Overview of the TREC 2014 Session Track*. Technical Report. TREC.
- [7] Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. QuAC: Question Answering in Context. In *EMNLP*. ACM, 2174–2184.
- [8] Gordon V. Cormack, Charles L A Clarke, and Stefan Buettcher. 2009. Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods. In *SIGIR*. ACM, 758.
- [9] W. Bruce Croft and R. H. Thompson. 1987. I3R: A New Approach to The Design of Document Retrieval Systems. *JASIST* 38, 6 (1987), 389–404.
- [10] J. Shane Culpepper, Fernando Diaz, and Mark D. Smucker. 2018. Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *SIGIR Forum* 52, 1 (2018), 34–90.
- [11] Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. 2019. CAsT 2019: The Conversational Assistance Track Overview. In *TREC 2019*. NIST.
- [12] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W. Bruce Croft. 2017. Neural Ranking Models with Weak Supervision. In *SIGIR*. ACM, 65–74.
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*. ACL, 4171–4186.
- [14] Laura Dietz, Manisha Verma, Filip Radlinski, and Nick Craswell. 2017. TREC Complex Answer Retrieval Overview. In *TREC*. NIST.
- [15] Ahmed Elgohary, Denis Peskov, and Jordan Boyd-Graber. 2019. Can You Unpack That? Learning to Rewrite Questions-in-Context. In *EMNLP*. ACM, 5920–5926.
- [16] Marziel Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data Augmentation for Low-Resource Neural Machine Translation. In *ACL*. ACL, 567–573.
- [17] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In *ACL*. ACL, 2–7.
- [18] Dongyi Guan, Hui Yang, and Nazli Goharian. 2012. Effective Structured Query Formulation for Session Search. In *TREC*. NIST.
- [19] Thorsten Joachims. 2002. Optimizing search Engines using Clickthrough Data. In *KDD*. ACM Press, 133.
- [20] Mandler Joshi, Omer Levy, Daniel S. Weld, and Luke Zettlemoyer. 2019. BERT for Coreference Resolution: Baselines and Analysis. *arXiv preprint arXiv:1908.09091* (2019).
- [21] Vineet Kumar and Sachindra Joshi. 2017. Incomplete Follow-up Question Resolution Using Retrieval Based Sequence to Sequence Learning. In *SIGIR*. ACM, 705–714.
- [22] Victor Lavrenko and W. Bruce Croft. 2001. Relevance Based Language Models. In *SIGIR*. ACM, 120–127.
- [23] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A Diversity-Promoting Objective Function for Neural Conversation Models. In *NAACL-HLT*. ACL, 110–119.
- [24] Yongjie Lin, Yi Chern Tan, and Robert Frank. 2019. Open Sesame: Getting Inside BERT’s Linguistic Knowledge. *arXiv preprint arXiv:1906.01698* (2019).
- [25] Sean MacAvaney, Andrew Yates, Arman Cohan, and Nazli Goharian. 2019. CEDR: Contextualized Embeddings for Document Ranking. In *SIGIR*. ACM, 1101–1104.
- [26] Mike Mintz, Steven Bills, Ron Snow, and Dan Jurafsky. 2009. Distant Supervision for Relation Extraction without Labeled Data. In *ACL*. ACL, 1003–1011.
- [27] Mandar Mitra, Amit Singhal, and Chris Buckley. 1998. Improving Automatic Query Expansion. In *SIGIR*. ACM, 206–214.
- [28] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset. *arXiv preprint arXiv:1611.09268* (2016).
- [29] Rodrigo Nogueira and Kyunghyun Cho. 2017. Task-Oriented Query Reformulation with Reinforcement Learning. In *EMNLP*. ACL, 574–583.
- [30] Robert N Oddy. 1977. Information Retrieval through Man-machine Dialogue. *Journal of Documentation* 33, 1 (1977), 1–14.
- [31] Kezban Dilek Onal, Ye Zhang, Ismail Sengor Altingovde, Md Mustafizur Rahman, Pinar Karagoz, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, Maarten de Rijke, and Matthew Lease. 2018. Neural Information Retrieval: At the End of the Early Years. *Information Retrieval Journal* 21, 2–3 (June 2018), 111–182.
- [32] Baolin Peng, Xiujun Li, Jianfeng Gao, Jingjing Liu, and Kam-Fai Wong. 2018. Deep Dyna-Q: Integrating Planning for Task-Completion Dialogue Policy Learning. In *ACL*. ACL, 2182–2192.
- [33] Yifan Qiao, Chenyan Xiong, Zhenghao Liu, and Zhiyuan Liu. 2019. Understanding the Behaviors of BERT in Ranking. *arXiv preprint arXiv:1904.07531* (2019).
- [34] Chen Qu, Liu Yang, Minghui Qiu, Yongfeng Zhang, Cen Chen, W. Bruce Croft, and Mohit Iyyer. 2019. Attentive History Selection for Conversational Question Answering. In *CIKM*. ACM, 1391–1400.
- [35] Filip Radlinski and Nick Craswell. 2017. A Theoretical Framework for Conversational Search. In *CIIR*. ACM, 117–126.
- [36] Dinesh Raghu, Sathish Indurthi, Jitendra Ajmera, and Sachindra Joshi. 2015. A statistical approach for Non-Sentential Utterance Resolution for Interactive QA System. In *SIGDIAL*. ACL, 335–343.
- [37] Siva Reddy, Danqi Chen, and Christopher D. Manning. 2019. CoQA: A Conversational Question Answering Challenge. *TACL* 7 (2019), 249–266.
- [38] Pengjie Ren, Zhumin Chen, Christof Monz, Jun Ma, and Maarten de Rijke. 2020. Thinking Globally, Acting Locally: Distantly Supervised Global-to-local Knowledge Selection for Background based Conversation. In *AAAI*.
- [39] Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. 2011. Named Entity Recognition in Tweets: An Experimental Study. In *EMNLP*. ACL, 1524–1534.
- [40] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *ACL*. ACL, 1073–1083.
- [41] Marc Sloan, Hui Yang, and Jun Wang. 2015. A Term-based Methodology for Query Reformulation Understanding. *Information Retrieval* 18, 2 (2015), 145–165.
- [42] Christophe Van Gysel, Evangelos Kanoulas, and Maarten de Rijke. 2016. Lexical Query Modeling in Session Search. In *ICTIR*. ACM, 69–72.
- [43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NIPS*. Curran Associates, Inc., 5998–6008.
- [44] Nikos Voskarides, Edgar Meij, Ridho Reinanda, Abhinav Khatan, Miles Osborne, Giorgio Stefanoni, Kambadar Prabhanjan, and Maarten de Rijke. 2018. Weakly-supervised Contextualization of Knowledge Graph Facts. In *SIGIR*. ACM, 765–774.
- [45] Alexandre Vtyurina, Denis Savenkov, Eugene Agichtein, and Charles L. A. Clarke. 2017. Exploring Conversational Search With Humans, Assistants, and Wizards. In *CHI*. ACM, 2187–2193.
- [46] Lidan Wang, Jimmy Lin, and Donald Metzler. 2011. A Cascade Ranking Model for Efficient Ranked Retrieval. In *SIGIR*. ACM, 105–114.
- [47] Hui Yang, Dongyi Guan, and Sicong Zhang. 2015. The Query Change Model: Modeling Session Search as a Markov Decision Process. *ACM Transactions on Information Systems* 33, 4 (2015), 20.
- [48] Wei Yang, Haotian Zhang, and Jimmy Lin. 2019. Simple Applications of BERT for Ad Hoc Document Retrieval. *arXiv preprint arXiv:1903.10972* (2019).
- [49] Yaosheng Yang, Wenliang Chen, Zhenghua Li, Zhengqiu He, and Min Zhang. 2018. Distantly Supervised NER with Partial Annotation Learning and Reinforcement Learning. In *COLING*. ACL, 2159–2169.
- [50] Mark Yatskar. 2019. A Qualitative Comparison of CoQA, SQuAD 2.0 and QuAC. In *NAACL-HLT*. ACL, 2318–2323.
- [51] Chengxiang Zhai and John Lafferty. 2001. A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval. In *SIGIR*. ACM, 334–342.