



















the Youth Innovation Promotion Association CAS under Grants No. 20144310, and 2016102, the National Key RD Program of China under Grants No. 2016QY02D0405, the Lenovo-CAS Joint Lab Youth Scientist Project, the K.C.Wong Education Foundation, and the Foundation and Frontier Research Key Program of Chongqing Science and Technology Commission (No. cstc2017jcyjBX0059).

## REFERENCES

- [1] Avi Arampatzis and Jaap Kamps. 2008. A Study of Query Length. In *Proceedings of the 31th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 811–812.
- [2] Nima Asadi, Donald Metzler, Tamer Elsayed, and Jimmy Lin. 2011. Pseudo Test Collections for Learning Web Search Ranking Functions. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 1073–1082.
- [3] Leif Azzopardi, Maarten de Rijke, and Krisztian Balog. 2007. Building Simulated Queries for Known-Item Topics: An Analysis Using Six European Languages. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 455–462.
- [4] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A Neural Probabilistic Language Model. *JMLR*, 1137–1155.
- [5] Wei-Cheng Chang, Felix X Yu, Yin-Wen Chang, Yiming Yang, and Sanjiv Kumar. 2019. Pre-training Tasks for Embedding-based Large-scale Retrieval. In *ICLR*.
- [6] Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training Text Encoders as Discriminators rather than Generators. In *ICLR*.
- [7] Prem C Consul and Gaurav C Jain. 1973. A Generalization of the Poisson Distribution. In *Technometrics*, Vol. 15. Taylor & Francis, 791–799.
- [8] Gordon V Cormack, Mark D Smucker, and Charles LA Clarke. 2011. Efficient and Effective Spam Filtering and Re-ranking for Large Web Datasets. In *Information retrieval*, Vol. 14. Springer, 441–465.
- [9] Zhuyun Dai and Jamie Callan. 2019. Deeper Text Understanding for IR with Contextual Neural Language Modeling. In *Proceedings of the 42th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 985–988.
- [10] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W Bruce Croft. 2017. Neural Ranking Models with Weak Supervision. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 65–74.
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, USA, 4171–4186.
- [12] Yixing Fan, Jiafeng Guo, Yanyan Lan, Jun Xu, Chengxiang Zhai, and Xueqi Cheng. 2018. Modeling Diverse Relevance Patterns in Ad-hoc Retrieval. In *The 41th international ACM SIGIR conference on research & development in information retrieval*. ACM, New York, NY, USA, 375–384.
- [13] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. A Deep Relevance Matching Model for Ad-hoc Retrieval. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 55–64.
- [14] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. Semantic Matching by Non-linear Word Transportation for Information Retrieval. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 701–710.
- [15] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. ACL, Stroudsburg, PA, USA, 328–339.
- [16] Kai Hui, Andrew Yates, Klaus Berberich, and Gerard de Melo. 2017. PACRR: A Position-Aware Neural IR Model for Relevance Matching. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, USA, 1049–1058.
- [17] Samuel Huston and W Bruce Croft. 2014. Parameters Learned in the Comparison of Retrieval Models using Term Dependencies. In *IR, UMASS*. Citeseer.
- [18] Kenton Lee, Ming-Wei Chang, and Kristina Toutanova. 2019. Latent Retrieval for Weakly Supervised Open Domain Question Answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. ACL, Stroudsburg, PA, USA, 6086–6096.
- [19] Xiaoyong Liu and W Bruce Croft. 2005. Statistical Language Modeling for Information Retrieval. In *IR, UMASS*.
- [20] Sean MacAvaney, Andrew Yates, Arman Cohan, and Nazli Goharian. 2019. CEDR: Contextualized Embeddings for Document Ranking. In *Proceedings of the 42th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 1101–1104.
- [21] Sean MacAvaney, Andrew Yates, Kai Hui, and Ophir Frieder. 2019. Content-Based Weak Supervision for Ad-Hoc Re-Ranking. In *Proceedings of the 42th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 993–996.
- [22] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Front-matter*. Cambridge University Press, i–iv.
- [23] Rodrigo Nogueira and Kyunghyun Cho. 2019. Passage Re-ranking with BERT. *arXiv preprint arXiv:1901.04085*.
- [24] Rodrigo Nogueira, Wei Yang, Kyunghyun Cho, and Jimmy Lin. 2019. Multi-stage Document Ranking with BERT. *arXiv preprint arXiv:1910.14424*.
- [25] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Jingfang Xu, and Xueqi Cheng. 2017. DeepRank: A New Deep Architecture for Relevance Ranking in Information Retrieval. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 257–266.
- [26] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, USA, 2227–2237.
- [27] Jay M Ponte and W Bruce Croft. 1998. A Language Modeling Approach to Information Retrieval. In *Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 275–281.
- [28] Tao Qin and Tie-Yan Liu. 2013. Introducing LETOR 4.0 datasets. *arXiv preprint arXiv:1306.2597*.
- [29] Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. Pre-trained Models for Natural Language Processing: A Survey. *arXiv preprint arXiv:2003.08271*.
- [30] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.
- [31] Stephen E Robertson and Steve Walker. 1994. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Springer, 232–241.
- [32] Erik F Sang and Fien De Mulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-independent Named Entity Recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. ACL, Stroudsburg, PA, USA, 142–147.
- [33] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1631–1642.
- [34] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2019. Mass: Masked Sequence to Sequence Pre-training for Language Generation. In *ICML*. 11328–11339.
- [35] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*. MIT Press, Cambridge, MA, USA, 3104–3112.
- [36] Wilson L Taylor. 1953. “Cloze Procedure”: A New Tool for Measuring Readability. In *Journalism quarterly*, Vol. 30. 415–433.
- [37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in neural information processing systems*. 5998–6008.
- [38] Wei Wang, Bin Bi, Ming Yan, Chen Wu, Zuyi Bao, Liwei Peng, and Luo Si. 2019. Structbert: Incorporating Language Structures into Pre-training for Deep Language Understanding. In *ICLR*.
- [39] Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-end Neural Ad-hoc Ranking with Kernel Pooling. In *Proceedings of the 40th International ACM SIGIR conference on research and development in information retrieval*. ACM, New York, NY, USA, 55–64.
- [40] Wei Yang, Haotian Zhang, and Jimmy Lin. 2019. Simple Applications of BERT for Ad Hoc Document Retrieval. *arXiv preprint arXiv:1903.10972* (2019).
- [41] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 5753–5763.
- [42] ChengXiang Zhai. 2008. Statistical Language Models for Information Retrieval. *Synthesis lectures on human language technologies* 1, 1 (2008), 1–141.
- [43] Chengxiang Zhai and John Lafferty. 2001. A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval. In *ACM SIGIR Forum*. New York, NY, USA, ACM, 268–276.
- [44] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with Extracted Gap-sentences for Abstractive Summarization. In *ICML*. 11328–11339.