

## References

- [1] Mohammad Aliannejadi, Julia Kiseleva, Aleksandr Chuklin, Jeff Dalton, and Mikhail Burtsev. 2021. Building and Evaluating Open-Domain Dialogue Corpora with Clarifying Questions. In *EMNLP*.
- [2] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (Paris, France) (SIGIR '19)*.
- [3] Reinald Kim Amplayo, Kellie Webster, Michael Collins, Dipanjan Das, and Shashi Narayan. 2023. Query Refinement Prompts for Closed-Book Long-Form QA. In *Annual Meeting of the Association for Computational Linguistics*. <https://api.semanticscholar.org/CorpusID:259370749>
- [4] Jaime Arguello, Adam Ferguson, Emery Fine, Bhaskar Mitra, Hamed Zamani, and Fernando Diaz. 2021. Tip of the tongue known-item retrieval: A case study in movie identification. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 5–14.
- [5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* abs/1409.0473 (2014). <https://api.semanticscholar.org/CorpusID:11212020>
- [6] N. J. Belkin, D. Kelly, G. Kim, J.-Y. Kim, H.-J. Lee, G. Muresan, M.-C. Tang, X.-J. Yuan, and C. Cool. 2003. Query length in interactive information retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Toronto, Canada) (SIGIR '03). Association for Computing Machinery, New York, NY, USA, 205–212. <https://doi.org/10.1145/860435.860474>
- [7] Ondrej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, et al. 2016. Findings of the 2016 conference on machine translation (wmt16). In *First conference on machine translation*. Association for Computational Linguistics, 131–198.
- [8] Paolo Boldi, Francesco Bonchi, Carlos Castillo, and Sebastiano Vigna. 2011. Query reformulation mining: models, patterns, and applications. *Inf. Retr.* 14, 3 (June 2011), 257–289. <https://doi.org/10.1007/s10791-010-9155-3>
- [9] Arthur Cámaras, David Maxwell, and Claudia Hauff. 2022. Searching, Learning, and Subtopic Ordering: A Simulation-Based Analysis. In *Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10–14, 2022, Proceedings, Part I* (Stavanger, Norway). Springer-Verlag, Berlin, Heidelberg, 142–156. [https://doi.org/10.1007/978-3-030-99736-6\\_10](https://doi.org/10.1007/978-3-030-99736-6_10)
- [10] Daniel Fernando Campos, Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, Li Deng, and Bhaskar Mitra. 2016. MS MARCO: A Human Generated MAchine REading COmprehension Dataset. *ArXiv* abs/1611.09268 (2016). <https://api.semanticscholar.org/CorpusID:1289517>
- [11] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. 2008. Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Las Vegas, Nevada, USA) (KDD '08). Association for Computing Machinery, New York, NY, USA, 875–883. <https://doi.org/10.1145/1401890.1401995>
- [12] Claudio Carpineto and Giovanni Romano. 2012. A Survey of Automatic Query Expansion in Information Retrieval. *ACM Comput. Surv.* 44, 1, Article 1 (jan 2012), 50 pages. <https://doi.org/10.1145/2071389.2071390>
- [13] Paul Alexandru Chirita, Claudiu S. Firan, and Wolfgang Nejdl. 2007. Personalized query expansion for the web (SIGIR '07). Association for Computing Machinery, New York, NY, USA, 7–14. <https://doi.org/10.1145/1277741.1277746>
- [14] Aleksandr Chuklin, Pavel Serdyukov, and Maarten de Rijke. 2013. Modeling clicks beyond the first result page. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 1217–1220.
- [15] Charles LA Clarke, Nick Craswell, and Ian Soboroff. 2009. Overview of the TREC 2009 Web Track. In *TREC*. Vol. 9. 20–29.
- [16] Charles LA Clarke, Nick Craswell, Ian Soboroff, and Ellen M Voorhees. 2011. Overview of the TREC 2011 Web Track.. In *TREC*.
- [17] Charles L. A. Clarke, Nick Craswell, Ian Soboroff, and Gordon V. Cormack. 2010. Overview of the TREC 2010 Web Track. In *Text Retrieval Conference*. <https://api.semanticscholar.org/CorpusID:16213318>
- [18] Charles L. A. Clarke, Nick Craswell, and Ellen M. Voorhees. 2012. Overview of the TREC 2012 Web Track. In *Text Retrieval Conference*. <https://api.semanticscholar.org/CorpusID:11517775>
- [19] Kevyn Collins-Thompson, Paul N. Bennett, Fernando Diaz, Charles L. A. Clarke, and Ellen M. Voorhees. 2013. TREC 2013 Web Track Overview. In *Text Retrieval Conference*. <https://api.semanticscholar.org/CorpusID:2180933>
- [20] Kevyn Collins-Thompson, Craig Macdonald, Paul N Bennett, Fernando Diaz, and Ellen M Voorhees. 2014. TREC 2014 Web Track Overview.. In *TREC*, Vol. 13. 1–15.
- [21] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, and Daniel Campos. 2020. Overview of the TREC 2020 deep learning track. In *TREC*.
- [22] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, Daniel Campos, and Ellen Voorhees. 2019. Overview of the TREC 2019 deep learning track. In *TREC 2019*.
- [23] Silviu Cucerzan and Ryan W White. 2007. Query suggestion based on user landing pages. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 875–876.
- [24] Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 2023. Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 10602–10621. <https://doi.org/10.18653/v1/2023.findings-emnlp.711>
- [25] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783* (2024).
- [26] Georges E Dupret and Benjamin Piwowarski. 2008. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 331–338.
- [27] Wieland Eckert, Esther Levin, and Roberto Pieraccini. 1997. User modeling for spoken dialogue system evaluation. In *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*. IEEE, 80–87.
- [28] Pierre Erbacher, Ludovic Denoyer, and Laure Soulard. 2022. Interactive query clarification and refinement via user simulation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2420–2425.
- [29] Pierre ERBACHER, Jian-Yun Nie, Philippe Preux, and Laure Soulard. 2024. Augmenting Ad-Hoc IR Dataset for Interactive Conversational Search. *Transactions on Machine Learning Research* (2024). <https://openreview.net/forum?id=z8d7nT1HWw>
- [30] Meiqi Guo, Mingda Zhang, Siva Reddy, and Malihe Alikhani. 2021. Abg-CoQA: Clarifying Ambiguity in Conversational Question Answering. In *3rd Conference on Automated Knowledge Base Construction*. <https://openreview.net/forum?id=SIDZ1o8FsJU>
- [31] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. 2009. Patterns of query reformulation during Web searching. *J. Am. Soc. Inf. Sci. Technol.* 60, 7 (July 2009), 1358–1371.
- [32] Vladimir Karpukhin, Barlas Ouzi, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- [33] Robert Krovetz. 1997. Homonymy and polysemy in information retrieval. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. 72–79.
- [34] Saar Kuzy, Anna Shtok, and Oren Kurland. 2016. Query Expansion Using Word Embeddings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (Indianapolis, Indiana, USA) (CIKM '16). Association for Computing Machinery, New York, NY, USA, 1929–1932. <https://doi.org/10.1145/2983323.2983876>
- [35] Shalom Lappin (Ed.). 1996. *The Handbook of Contemporary Semantic Theory*. Blackwell Reference, Cambridge, Mass., USA. 329–423 pages.
- [36] Tessa Lau and Eric Horvitz. 1999. Patterns of search: analyzing and modeling Web query refinement. In *Proceedings of the Seventh International Conference on User Modeling* (Banff, Canada) (UM '99). Springer-Verlag, Berlin, Heidelberg, 119–128.
- [37] Dongryeo Lee, Segwang Kim, Minwoo Lee, Hwanhee Lee, Joonsuk Park, Sang-Woo Lee, and Kyomin Jung. 2023. Asking Clarification Questions to Handle Ambiguity in Open-Domain QA. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 11526–11544. <https://doi.org/10.18653/v1/2023.findings-emnlp.772>
- [38] Sean Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 7871–7880. <https://doi.org/10.18653/v1/2020.acl-main.703>
- [39] Sean MacAvaney, Andrew Yates, Sergey Feldman, Doug Downey, Arman Cohan, and Nazli Goharian. 2021. Simplified Data Wrangling with ir\_datasets. In *SIGIR*.
- [40] Iain Mackie, Jeffrey Dalton, and Andrew Yates. 2021. How deep is your learning: The DL-HARD annotated deep learning dataset. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2335–2341.
- [41] David Maxwell, Leif Azzopardi, Kalervo Järvelin, and Heikki Keskustalo. 2015. Searching and stopping: An analysis of stopping rules and strategies. In *Proceedings of the 24th ACM international on conference on information and knowledge management*. 313–322.

- [42] Qiaozhu Mei, Dengyong Zhou, and Kenneth Church. 2008. Query suggestion using hitting time. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (Napa Valley, California, USA) (CIKM '08). Association for Computing Machinery, New York, NY, USA, 469–478. <https://doi.org/10.1145/1458082.1458145>
- [43] Sewon Min, Julian Michael, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2020. AmbigQA: Answering Ambiguous Open-domain Questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 5783–5797. <https://doi.org/10.18653/v1/2020.emnlp-main.466>
- [44] Dan Morris, Meredith Ringel Morris, and Gina Venolia. 2008. SearchBar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 1207–1216. <https://doi.org/10.1145/1357054.1357242>
- [45] Rodrigo Nogueira and Kyunghyun Cho. 2019. Passage Re-ranking with BERT. *arXiv preprint arXiv:1901.04085* (2019).
- [46] Rodrigo Nogueira, Zhiying Jiang, Ronak Pradeep, and Jimmy Lin. 2020. Document Ranking with a Pretrained Sequence-to-Sequence Model. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 708–718. <https://doi.org/10.18653/v1/2020.findings-emnlp.63>
- [47] Paul Owoicho, Ivan Sekulic, Mohammad Aliannejadi, Jeffrey Dalton, and Fabio Crestani. 2023. Exploiting Simulated User Feedback for Conversational Search: Ranking, Rewriting, and Beyond. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) (SIGIR '23). Association for Computing Machinery, New York, NY, USA, 632–642. <https://doi.org/10.1145/3539618.3591683>
- [48] Dragomir R. Radev, Hong Qi, Harris Wu, and Weiguo Fan. 2002. Evaluating Web-based Question Answering Systems. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Manuel González Rodríguez and Carmen Paz Suarez Araujo (Eds.). European Language Resources Association (ELRA), Las Palmas, Canary Islands - Spain. <http://www.lrec-conf.org/proceedings/lrec2002/pdf/301.pdf>
- [49] Filip Radlinski and Nick Craswell. 2017. A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) (CHIIR '17). Association for Computing Machinery, New York, NY, USA, 117–126. <https://doi.org/10.1145/3020165.3020183>
- [50] Hossein A. Rahmani, Xi Wang, Yue Feng, Qiang Zhang, Emine Yilmaz, and Aldo Lipani. 2023. A Survey on Asking Clarification Questions Datasets in Conversational Systems. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Anna Rogers, Jordan Boyd-Graber, and Naoki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 2698–2716. <https://doi.org/10.18653/v1/2023.acl-long.152>
- [51] Sudha Rao and Hal Daumé III. 2018. Learning to Ask Good Questions: Ranking Clarification Questions using Neural Expected Value of Perfect Information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, Melbourne, Australia, 2737–2746. <https://doi.org/10.18653/v1/P18-1255>
- [52] Ivan Sekulić, Mohammad Aliannejadi, and Fabio Crestani. 2021. Towards Facet-Driven Generation of Clarifying Questions for Conversational Search. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval* (Virtual Event, Canada) (ICTIR '21). Association for Computing Machinery, New York, NY, USA, 167–175. <https://doi.org/10.1145/3471158.3472257>
- [53] 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 - Volume 2* (Montreal, Canada) (NIPS'14). MIT Press, Cambridge, MA, USA, 3104–3112.
- [54] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Huai hsin Chi, and Denny Zhou. 2022. Self-Consistency Improves Chain of Thought Reasoning in Language Models. *ArXiv* abs/2203.11171 (2022). <https://api.semanticscholar.org/CorpusID:247595263>
- [55] Yining Wang, Liwei Wang, Yuanzhi Li, Di He, and Tie-Yan Liu. 2013. A Theoretical Analysis of NDCG Type Ranking Measures. In *Annual Conference Computational Learning Theory*. <https://api.semanticscholar.org/CorpusID:3804244>
- [56] Zhenduo Wang, Yuancheng Tu, Corby Rosset, Nick Craswell, Ming Wu, and Qingyan Ai. 2023. Zero-shot Clarifying Question Generation for Conversational Search. In *Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) (WWW '23). Association for Computing Machinery, New York, NY, USA, 3288–3298. <https://doi.org/10.1145/3543507.3583420>
- [57] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 24824–24837. [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf)
- [58] Jingjing Xu, Yuechen Wang, Duyu Tang, Nan Duan, Pengcheng Yang, Qi Zeng, Ming Zhou, and Xu Sun. 2019. Asking Clarification Questions in Knowledge-Based Question Answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 1618–1629. <https://doi.org/10.18653/v1/D19-1172>
- [59] Hamed Zamani, Susan Dumais, Nick Craswell, Paul Bennett, and Gord Lueck. 2020. Generating clarifying questions for information retrieval. In *Proceedings of the web conference 2020*. 418–428.
- [60] Tianyi Zhang\*, Varsha Kishore\*, Felix Wu\*, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=SkeHuCVFDr>
- [61] Tong Zhang, Peixin Qin, Yang Deng, Chen Huang, Wenqiang Lei, Junhong Liu, Dingnan Jin, Hongru Liang, and Tat-Seng Chua. 2024. CLAMBER: A Benchmark of Identifying and Clarifying Ambiguous Information Needs in Large Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikanth (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 10746–10766. <https://doi.org/10.18653/v1/2024.acl-long.578>
- [62] Yujia Zhou, Jing Yao, Zhicheng Dou, Yiteng Tu, Ledell Wu, Tat-Seng Chua, and Ji-Rong Wen. 2024. ROGER: Ranking-oriented Generative Retrieval. *ACM Trans. Inf. Syst.* (June 2024). <https://doi.org/10.1145/3603167> Just Accepted.
- [63] Jie Zou, Mohammad Aliannejadi, Evangelos Kanoulas, Maria Soledad Pera, and Yiqun Liu. 2023. Users Meet Clarifying Questions: Toward a Better Understanding of User Interactions for Search Clarification. *ACM Trans. Inf. Syst.* 41, 1, Article 16 (jan 2023), 25 pages. <https://doi.org/10.1145/3524110>