

# WaterlooClarke at the TREC 2022 Conversational Assistant Track

Siqing Huo, Xinyi Yan, Charles L. A. Clarke

October 31, 2022

## 1 Introduction

The WaterlooClarke group submitted three runs to TREC Conversational Assistant Track (CAST) 2022:

- UWCmanual22
- UWCauto22
- UWCcano22

This report describes the techniques used and the results of these three runs. The pipeline to generate each run contains three main steps: 1) query reformulation, 2) passage retrieval and reranking, 3) passage summarization and reranking. For the UWCmanual22 run, we used the manual rewritten utterances provided as reformulated queries. For the UWCauto22 run, we used the automatic rewritten utterances provided as reformulated queries. For the UWCcano22 run, we augmented the automatic rewritten utterances provided with keywords extracted from the previous responses' titles, and used them as reformulated queries. We continued last year's strategy [6] of performing passage reranking with both MonoT5 and DuoT5 [2] on the pooled retrieved passages from both sparse and dense retrievers [1]. This year we focused on experimenting with the following changes: 1) Incorporating generative summarization techniques to make the produced answers more conversational. 2) Incorporating pseudo-relevance feedback into the passage retrieval stage to improve performance. In the next section, we will present the details of our pipeline.

## 2 Methodology

### 2.1 Query Reformulation

Similar to previous years, raw utterances are informal and often omit details that need to be inferred from the context. Therefore, query reformulation is a crucial step, in which essential information from the context is added to the raw utterances. This is the only part of the pipeline where each run uses a different approach.

The UWCmanual22 run uses provided manual rewritten utterances, where human manually expand the raw utterances with all the necessary information from the conversation history.

The UWCauto22 run uses provided automatically rewritten utterances, where a finetuned T5 model ChattyGoose [8] is used to automatically rewrite the query using conversation history. The conversation history consists of all the raw utterances and responses before the current query that is to be rewritten. Let  $u_i$  denote the  $i^{th}$  raw utterance, and  $r_i$  denote the response to  $u_i$ , the conversation history for  $u_i$  is  $[u_1, r_1, \dots, u_{i-1}, r_{i-1}]$ .

The UWCcano22 run augments provided automatically rewritten utterances, by appending keywords from the previous automatically rewritten utterance, response passage, and title of the response passage. Let  $a_i$  denote the automatically rewritten output of  $u_i$ ,  $a_i$  is augmented with keywords of  $a_{i-1}$ ,  $r_{i-1}$ , and the title of the raw passage corresponds to  $r_{i-1}$ . To extract the keywords, the RAKE algorithm [4] is used.

	UWCmanual22				UWCauto22				UWCcano22			
	lenient		strict		lenient		strict		lenient		strict	
	map	ndcg cut 20	map	ndcg cut 20	map	ndcg cut 20	map	ndcg cut 20	map	ndcg cut 20	map	ndcg cut 20
# <median	20	21	20	21	60	61	64	61	45	47	45	47
# = median	0	0	0	0	0	1	0	1	1	0	0	0
# >median	122	120	114	120	92	93	90	93	107	109	110	109
# = max	21	22	29	22	11	8	9	8	10	7	8	7

Table 1: TREC CAsT 2022 Preliminary Results.

## 2.2 Passage Retrieval and Reranking

Similar to last year, we used the BERT-based dense retriever ANCE [5] and a BM25-based retriever finetuned to maximize recall@1000 over previous years’ TREC CAsT data. Also similar to last year, we pooled the retrieved results from both retrievers, and then applied MonoT5 and DuoT5 [2] to rerank the retrieved pool of passages.

In addition, we incorporated pseudo-relevance feedback (PRF) for the dense retriever ANCE as well. Recent study [7] demonstrates that ANCE-PRF can lead to significant accuracy improvements. We combine the original query with the top three reranked passages to obtain a new query embedding, and then used ANCE to do another pass of passage retrieval with the new query embedding. The newly retrieved passages are merged into the previous pool of passages, and another pass of passage reranking with MonoT5 and DuoT5 is performed to obtain the final results.

## 2.3 Passage Summarization and Reranking

This year we experimented with generative passage summarization using T5 [3]. Passage summarization can reduce the amount of extraneous information in the returned passages, which makes the outputs more like a real answer compared to a raw passage. We take the top twenty reranked passages from the previous step, summarize them and feed the summarized passages into MonoT5 and DuoT5 to obtain the final ranked list of responses.

## 3 Results

We compared our results with the released performance benchmarks this year. The benchmark including the median and maximum score across all participating teams this year. Our performance is summarized in Table 1. As expected, using the manually rewritten utterances yield the best performance, which implies more work could be done to improve the query reformulation stage. Through comparing the results for UWCauto22 and UWCcano22, it seems augmenting automatically rewritten utterances with extracted keywords help improving the performance for those originally performing badly, but also brings down the performance of some that originally performs very well. This suggests that our attempt of augmenting automatically rewritten utterances is successful in general, but sometimes the extracted keywords may introduce extraneous or irrelevant information. This suggests we can experiment with better topic detection methods to further improve the performance.

## 4 Conclusion

In TREC CAsT 2022, our team further improved our passage retrieval by performing ANCE-PRF on top reranked passages. We also experimented with incorporating generative passage summarization into our pipeline. In the future, we can experiment with query-based summarization methods, finetuning the contextual query rewriting model, and augment automatically rewritten queries using better topic detection methods. We look forward to participating in TREC CAsT 2023.

## References

- [1] Negar Arabzadeh, Xinyi Yan, and Charles L. A. Clarke. “Predicting Efficiency/Effectiveness Trade-offs for Dense vs. Sparse Retrieval Strategy Selection”. In: *CoRR* abs/2109.10739 (2021). arXiv: [2109.10739](https://arxiv.org/abs/2109.10739). URL: <https://arxiv.org/abs/2109.10739>.

- [2] Ronak Pradeep, Rodrigo Nogueira, and Jimmy Lin. “The Expando-Mono-Duo Design Pattern for Text Ranking with Pretrained Sequence-to-Sequence Models”. In: *CoRR* abs/2101.05667 (2021). arXiv: [2101.05667](https://arxiv.org/abs/2101.05667). URL: <https://arxiv.org/abs/2101.05667>.
- [3] Colin Raffel et al. “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”. In: *CoRR* abs/1910.10683 (2019). arXiv: [1910.10683](https://arxiv.org/abs/1910.10683). URL: <http://arxiv.org/abs/1910.10683>.
- [4] Stuart Rose et al. “Automatic Keyword Extraction from Individual Documents”. In: Mar. 2010, pp. 1–20. ISBN: 9780470689646. DOI: [10.1002/9780470689646.ch1](https://doi.org/10.1002/9780470689646.ch1).
- [5] Lee Xiong et al. “Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval”. In: *CoRR* abs/2007.00808 (2020). arXiv: [2007.00808](https://arxiv.org/abs/2007.00808). URL: <https://arxiv.org/abs/2007.00808>.
- [6] Xinyi Yan, Charles L. A. Clarke, and Negar Arabzadeh. *WaterlooClarke at the TREC 2021 Conversational Assistant Track*. URL: <https://trec.nist.gov/pubs/trec30/papers/WaterlooClarke-CAsT.pdf>.
- [7] HongChien Yu, Chenyan Xiong, and Jamie Callan. “Improving Query Representations for Dense Retrieval with Pseudo Relevance Feedback”. In: *CoRR* abs/2108.13454 (2021). arXiv: [2108.13454](https://arxiv.org/abs/2108.13454). URL: <https://arxiv.org/abs/2108.13454>.
- [8] Edwin Zhang et al. “Chatty Goose: A Python Framework for Conversational Search”. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’21. Virtual Event, Canada: Association for Computing Machinery, 2021, pp. 2521–2525. ISBN: 9781450380379. DOI: [10.1145/3404835.3462782](https://doi.org/10.1145/3404835.3462782). URL: <https://doi.org/10.1145/3404835.3462782>.