Improving Query Representations for Dense Retrieval with Pseudo Relevance Feedback

HongChien Yu  
Carnegie Mellon University  
hongqiay@cs.cmu.edu

Chenyan Xiong  
Microsoft Research  
chenyan.xiong@microsoft.com

Jamie Callan  
Carnegie Mellon University  
callan@cs.cmu.edu

ABSTRACT

Dense retrieval systems conduct first-stage retrieval using embedded representations and simple similarity metrics to match a query to documents. Its effectiveness depends on encoded embeddings to capture the semantics of queries and documents, a challenging task due to the shortness and ambiguity of search queries. This paper proposes ANCE-PRF, a new query encoder that uses pseudo relevance feedback (PRF) to improve query representations for dense retrieval. ANCE-PRF uses a BERT encoder that consumes the query and the top retrieved documents from a dense retrieval model, ANCE, and it learns to produce better query embeddings directly from relevance labels. It also keeps the document index unchanged to reduce overhead. ANCE-PRF significantly outperforms ANCE and other recent dense retrieval systems on several datasets. Analysis shows that the PRF encoder effectively captures the relevant and complementary information from PRF documents, while ignoring the noise with its learned attention mechanism.

KEYWORDS

• Information systems → Retrieval models and ranking

CCS CONCEPTS

- Information systems → Retrieval models and ranking

ACM Reference Format:

HongChien Yu, Chenyan Xiong, and Jamie Callan. 2021. Improving Query Representations for Dense Retrieval with Pseudo Relevance Feedback. In Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM ’21), November 1–5, 2021, Virtual Event, QLD, Australia. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459637.3482124

1 INTRODUCTION

Dense retrieval systems first encode queries and documents into a dense embedding space and then perform information retrieval by finding a query’s nearest neighbors in the embedding space [16, 20, 27, 33]. With the advancement of pre-trained language models [8, 23], dedicated training strategies [16, 33], and efficient nearest neighbor search [12, 15], dense retrieval systems have shown effectiveness in a wide range of tasks, including web search [26], open domain question answering [18], and zero-shot IR [29].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CIKM ’21, November 1–5, 2021, Virtual Event, QLD, Australia
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8446-9/21/11
https://doi.org/10.1145/3459637.3482124

Retrieval with dense, fully-learned representations has the potential to address some fundamental challenges in sparse retrieval. For example, vocabulary mismatch can be solved if the embeddings accurately capture the information need behind a query and maps it to relevant documents. However, decades of IR research demonstrates that inferring a user’s search intent from a concise and often ambiguous search query is challenging [7]. Even with powerful pre-trained language models, it is unrealistic to expect an encoder to perfectly embed the underlying information need from a few query terms.

A common technique to improve query understanding in sparse retrieval systems is pseudo relevance feedback (PRF) [7, 19, 34], which uses the top retrieved documents from an initial search as additional information to enrich the query representation. Whether PRF information is used via query expansion [14, 34] or query term reweighting [2], its efficacy has been consistently observed across various search scenarios, rendering PRF a standard practice in many sparse retrieval systems.

This work leverages PRF information to improve query representations in dense retrieval. Given the top retrieved documents from a dense retrieval model, e.g., ANCE [33], we build a PRF query encoder, ANCE-PRF, that uses a BERT encoder [8] to consume the query and the PRF documents to refine the query representation. ANCE-PRF is trained end-to-end using relevance labels and learns to optimize the query embeddings using the rich information from PRF documents. It reuses the document index from ANCE to avoid duplicating index storage.

In experiments on MS MARCO and TREC Deep Learning (DL) Track passage ranking benchmarks, ANCE-PRF is consistently more accurate than ANCE and several recent dense retrieval systems that use more sophisticated models and training strategies [24, 36]. We also observe large improvements on DL-HARD [25] queries, a curated set to include complex search intents challenging for neural systems. To the best of our knowledge, ANCE-PRF is among the best performing first-stage retrieval systems on the highly competitive MARCO passage ranking leaderboard.

Our studies confirm that the advantages of ANCE-PRF reside in its ability to leverage the useful information from the PRF documents while ignoring the noise from irrelevant PRF documents. The PRF encoder allocates substantially more attention to terms in the relevant PRF documents, compared to those from the irrelevant documents. A case study shows that the encoder focuses more on PRF terms that are complementary to the query terms in representing search intents. These help ANCE-PRF learn better query embeddings that are closer to the relevant documents and improve the majority of testing queries.

1Our code, checkpoints, and ranking results are open-sourced at https://github.com/yuhongqian/ANCE-PRF.
2 RELATED WORK

In dense retrieval systems, queries and documents are encoded by a dual-encoder, often BERT-based, into a shared embedding space [16, 17, 24, 33]. Recent research in dense retrieval mainly focuses on improving the training strategies, especially the negative sampling part, including random sampling in batch [33], sampling from BM25 top negatives [9, 20], sampling from an asynchronously [33] updated hard negatives index, constructing hard negatives using document index from an existing dense retrieval model [36], or reranking models [13, 22]. Most dense retrieval systems encode a document using a constant number of embedding vectors [16, 24, 33], often one per document. There are also approximation using its Transformer attention. Specifically, the attention index for PRF. Therefore, only

\[ \text{cls\_attention}(t_j) = \sum_{i} \frac{\exp(q_{cls} \cdot k_j)}{\sum_{i} \exp(q_{cls} \cdot k_i)}, \]

where \( q_{cls} \) and \( k_j \) are the “query” vector and \( j \)th input token’s “key” vector of the \( i \)th attention head [30]. Ideally, the PRF encoder should learn to yield

\[ \sum_{j} \text{cls\_attention}(t_j) > \sum_{j} \text{cls\_attention}(t_{j^*}) \]

where \( j^* \) are indexes of the meaningful tokens from the PRF documents, and \( j \) are those of the irrelevant PRF tokens.

ANCE-PRF can be easily integrated with any dense retrieval models. With the document embeddings and index unchanged, the only computational overheads are one more query encoder forward pass (Eq. (2)) and one more nearest neighbor search (Eq. (3)), a minor addition to the dense retrieval process [33].

4 EXPERIMENTAL SETUP

Next, we discuss the datasets, baselines, and implementation details.

Datasets. We use MS MARCO passage training data [26] which includes 530K training queries. We first evaluate on its dev set with 7k queries and also obtain the testing results by submitting to its leaderboard. MARCO’s official metric is MRR@10.

We also evaluate the MARCO trained model on two additional evaluation benchmarks, TREC DL [5, 6] and DL-HARD [25]. TREC DL [5, 6] includes 43 labeled queries from 2019 and 54 from 2020 for the MARCO corpus. The official metric is NDCG@10 and Recall@1K, the latter with label binarized at relevance point 2. Following Xiong et al. [33], we also report HOLE@10, the unjudged fraction of top 10 retrieved documents, to reflect the coverage of pooled labels on dense retrieval systems. DL-HARD [25] contains 50 queries from TREC DL that were curated to challenge neural systems in a prior TREC DL track. Its official metric is NDCG@10.

Baselines include BM25 [28], RM3 [14, 19], a classical PRF framework in sparse retrieval. We also compare with several recent dense retrievers. ME-BERT [24] was trained with hard-negative mining [10], and is the only one that uses multi-vector document encoding. DE-BERT [24] is the single-vector version of ME-BERT. DPR [16] is trained with in-batch negatives. LTRe [36] generates hard negatives using document embeddings from an existing dense retrieval model. ANCE [33] uses hard negatives from asynchronously updated dense retrieval index using the latest model checkpoint.

Implementation Details. In training, we initialize query encoder from the ANCE FirstP model [33]2 and kept the document embeddings from ANCE (and thus also the ANCE negative index) unchanged. All hyperparameters used in ANCE training are inherited in ANCE-PRF. All models are trained on two RTX 2080 Ti GPUs with per-GPU batch size 4 and gradient accumulation step 8 for 450K steps. We keep the model checkpoint with the best MRR@10 score on the MS MARCO dev set.

In inference, we first obtain ANCE top \( k \) documents using Faiix INDEXFlatP and Eq. (1), feed them into the ANCE-PRF query encoder (Eq. (2)) for updated query embeddings, and run another Faiix search with Eq. (3) for final results.

2https://github.com/microsoft/ANCE
Table 1: Ranking results. ANCE-PRF uses 3 feedback documents. All baseline results except BM25 and BM25+RM3 are reported by previous work. Statistically significant improvements over baselines are indicated by * (BM25), † (BM25+RM3), ‡ (DE-BERT), § (ME-BERT), and ¶ (ANCE) with p ≤ 0.05 in t-test. Per query results of those underlined are not available for significance tests.

Table 2: Ranking accuracy with a varying number of PRF documents (k). Avg_Rel is the average relevance score of PRF documents at position k. Superscripts* mark statistically significant improvements over k. ANCE results are in the first row (*).

Table 3: Results on DL-HARD [25]. We use the same symbols as in Table 1 for statistically significant improvements.

5 EXPERIMENTAL RESULTS

In this section, we discuss our experimental results and studies.

5.1 Overall Results

Table 1 includes overall retrieval accuracy on MS MARCO and TREC DL datasets. ANCE-PRF outperforms ANCE, its base retrieval system, on all datasets. On the challenging DL-HARD (Table 3), ANCE-PRF improves NDCG@10 By 9.3% over ANCE, indicating ANCE-PRF’s advantage in queries challenging for neural systems. These results suggest that ANCE-PRF effectively leverages PRF information to produce better query embeddings. ANCE-PRF also helps retrieve relevant documents not recognized by ANCE, improving R@1K by about 5% on both TREC DL sets.

ANE PRF rankings are significantly more precise than the sparse retrieval baselines with large margins across all datasets. RM3 achieves the best R@1K on both TREC DL sets, but its improvement is not as significant on DL-HARD.

ANE PRF also outperforms several strong dense retrieval baselines and produces the most accurate rankings on almost all datasets. While Luan et al. [24] discuss the theoretical benefits of higher dimensional dense retrieval as in ME-BERT, our empirical results show that a well-informed query encoder can achieve comparable results, while avoiding the computational and spatial overhead caused by using multiple vectors per document.

5.2 Ablation on PRF Depths

To understand the number of feedback documents (k) needed for effective learning, we trained models using different k and report the results in Table 2. We trained k = 0 as a controlled experiment, which is equivalent to training ANCE for an extra 450K steps with fixed negatives.

Overall, we observe that models with k > 0 are consistently better than ANCE (k = +) and k = 0, showing that ANCE-PRF effectively utilizes the given feedback relevance information. The Avg_Rel indicates that PRF documents at k > 1 contain noisy relevance information, which is a known challenge for traditional PRF approaches [4]. Nevertheless, ANCE-PRF yields stable improvements over ANCE for k = 1 to 5, demonstrating the model’s robustness against noisy feedback from deeper k.

5.3 Analyses of Embedding Space & Attention

In this group of experiments, we analyze the learned embeddings and attention in ANCE-PRF.

Embedding Space. Fig. 1(a) shows the distance during training between the ANCE-PRF query embedding and the embeddings of the original ANCE query, the relevant documents, and the irrelevant documents. We use MARCO dev in this study, in which about one out of the three PRF documents is relevant. In the embedding space, ANCE-PRF queries are closest to the original query and then the relevant documents, while further away from the irrelevant documents. ANCE-PRF’s query embeddings effectively encode both the query and the feedback relevance information.

Learned Attention. We also analyze the learned attention on the relevant and the irrelevant PRF documents during training. We use TREC DL 2020 for this study as its dense relevance labels provide more stable observations. We calculate the average attention from the [CLS] token to each group (“relevant”, “irrelevant”, and “all”) of PRF document (Eq. (5) & (6)), and plot them in Fig. 1(b)-1(d).
Figure 1: Fig. (a) shows the dot product between the ANCE-PRF query embedding and document embeddings at different training steps (x-axis). Fig. (b)-(d) are the cls_attention (y-axes) on “all”, “relevant”, and “irrelevant” feedback documents ranked at positions 1-3 in the initial retrieval.

Figure 2: The histograms show the per-query NDCG@10 differences between ANCE-PRF and ANCE retrieval on TREC DL 2020’s 54 queries. ANCE-PRF wins on 34 queries, loses on 15, and ties on 5. The tables show the example queries and the two models’ first different retrieved passages. Terms receiving higher attention weights are highlighted in darker red.

As training proceeds, ANCE-PRF pays more and more attention to the relevant PRF documents than the irrelevant ones, showing the effectiveness of its learning. Note that the original query always attracts the most attention from the PRF encoder, which is intuitive, as the majority of the search intent is to be determined by the query. The PRF information is to refine the query representation with extra information but not to invalidate it.

5.4 Case Study
Fig. 2 plots the per query win/loss of ANCE-PRF versus ANCE on TREC DL 2020 and shows one example each.

ANCE-PRF wins on more queries and with larger margins. We also notice the PRF query encoder focuses more on terms that are complementary to the query. In the winning example, ANCE-PRF picks up terms explaining what “un fao” is and does not mistake “un” as “uno”. On the other hand, ANCE-PRF may be misled by information appearing in multiple feedback documents. This is a known challenge for PRF because the correctness of information from multiple feedback documents is its core assumption [19]. In the losing example, “pattern their shotguns” occurs in multiple PRF documents, attracting too much attention to allow ANCE-PRF to make a better choice.

6 CONCLUSION
Existing dense retrievers learn query representations from short and ambiguous user queries, thus a query representation may not precisely reflect the underlying information need. ANCE-PRF addresses this problem with a new query encoder that learns better query representations from the original query and the top-ranked documents from a state-of-the-art dense retriever, ANCE.

Our experiments demonstrate that ANCE-PRF’s effectiveness in refining query understanding and its robustness against noise from imperfect feedback. Our studies reveal that ANCE-PRF learns to distinguish between relevant and irrelevant documents. We show that ANCE-PRF successfully learns to identify relevance information with its attention mechanism. Its query encoder pays more attention to the relevant portion of the PRF documents, especially the PRF terms that complement the query terms in expressing the information need.

ANCE-PRF provides a straightforward way to leverage the PRF information in dense retrieval and can be used as a plug-in in embedding-based retrieval systems. We observe that simply leveraging the classic PRF information in the new neural-based retrieval regime leads to significant accuracy improvements, suggesting that more future research can be done in this direction.
REFERENCES

[1] Qingyao Ai, KePing Bi, Jiafeng Guo, and W. Bruce Croft. 2018. Learning a Deep Listwise Context Model for Ranking Refinement. In The 41st International ACM SIGIR Conference on Research Development in Information Retrieval. ACM, 135–144.

[2] Michael Bendersky, Donald Metzler, and W Bruce Croft. 2011. Parameterized concept weighting in verbose queries. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 605–614.

[3] Xiaoyang Chen, Kai Hui, Ben He, Xianpei Han, Le Sun, and Zheng Ye. 2021. Co-BERT: A Context-Aware BERT Retrieval Model Incorporating Local and Query-specific Context. arXiv preprint arXiv:2104.08323 (2021).

[4] Kevyn Collins-Thompson. 2009. Reducing the risk of query expansion via robust constrained optimization. In Proceedings of the 18th ACM Conference on Information and Knowledge Management. ACM, 857–866.

[5] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, and Daniel Campos. 2020. Overview of the TREC 2020 Deep Learning Track. In Proceedings of the Twenty-Ninth Text REtrieval Conference (NIST Special Publication). National Institute of Standards and Technology (NIST).

[6] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, Daniel Campos, and Ellen M. Voorhees. 2020. Overview of the TREC 2019 Deep Learning Track. In Proceedings of the Twenty-Eighth Text REtrieval Conference (NIST Special Publication). National Institute of Standards and Technology (NIST).

[7] W Bruce Croft, Donald Metzler, and Trevor Strohman. 2010. Parameterized Query Expansion via Concept Weighting. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 55–64.

[8] Xiaoyang Chen, Kai Hui, Ben He, Xianpei Han, Le Sun, and Zheng Ye. 2021. Co-BERT: A Context-Aware BERT Retrieval Model Incorporating Local and Query-specific Context. arXiv preprint arXiv:2104.08323 (2021).

[9] Luyu Gao, Zhuyun Dai, Tongfei Chen, Zhen Fan, Benjamin Van Durme, and Jacob Devlin. 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, Vol. 1. Association for Computational Linguistics, 417–4186.

[10] Luyu Gao, Zhuyun Dai, Tongfei Chen, Zhen Fan, Benjamin Van Durme, and Jacob Devlin. 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, Vol. 1. Association for Computational Linguistics, 417–4186.

[11] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur P. Parikh, Chris Alberti, Danielle Epstein, Ilia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. 2019. Natural Questions: a Benchmark for Question Answering Research. Trans. Assoc. Comput. Linguistics 7 (2019), 452–466.

[12] Victor Lavrenko and W. Bruce Croft. 2001. Relevance-Based Language Models. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 120–127.

[13] Kenton Lee, Ming-Wei Chang, and Kristina Toutanova. 2019. Latent Retrieval for Weakly Supervised Open Domain Question Answering. In Proceedings of the 57th Conference of the Association for Computational Linguistics, Vol. 1. Association for Computational Linguistics, 6086–6096.

[14] Canja Li, Yingfes Sun, Ben He, Le Wang, Kai Hui, Andrew Yates, Le Sun, and Jungxun Xu. 2018. NPFR: A Neural Pseudo Relevance Feedback Framework for Ad-hoc Information Retrieval. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 4482–4491.

[15] Sheng-Chieh Lin, Jheng-Hong Yang, and Jimmy Lin. 2020. Distilling Dense Representations for Ranking using Tightly-Coupled Teachers. arXiv preprint arXiv:2010.11386 (2020).

[16] Shuai Lu, Chenyuan Xiong, Di He, Guolin Ke, W. Bruce Croft, and Hongying Dou. 2021. FusedDense: A Deep Listwise Model for Dense Retrieval. arXiv preprint arXiv:2102.09206 (2021).

[17] Yi Luan, Jacob Eisenstein, Kristina Toutanova, and Michael Collins. 2021. Sparse, Dense, and Attentional Representations for Text Retrieval. Trans. Assoc. Comput. Linguistics 9 (2021), 329–345.

[18] Iain Mackie, Jeffrey Dalton, and Andrew Yates. 2021. How Deep is your Learning: the DL-HARD Annotated Deep Learning Dataset. In Proceedings of the 46th International ACM SIGIR conference on research and development in Information Retrieval. ACM, 2335–2341.

[19] Tri Nguyen, Mir Rosenberg, Xiao Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A Human Generated Machine Reading Comprehension Dataset. In Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (CEUR Workshop Proceedings, Vol. 1773). CEUR-WS.org.

[20] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 4512–4525.

[21] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 4512–4525.

[22] Sanjiv Kumar. 2020. Accelerating Large-Scale Inference with Anisotropic Vector Quantization. In Proceedings of the 37th International Conference on Machine Learning. PMLR, 3887–3896.

[23] 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 30: Annual Conference on Neural Information Processing Systems. 5998–6008.

[24] Xiao Wang, Craig Macdonald, Nicola Tomellotito, and Iadh Ounis. 2021. Pseudo-Relevance Feedback for Multiple Representation Dense Retrieval. arXiv preprint arXiv:2106.11251.

[25] Chenyan Xiong, Zhiyuan 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, 55–64.

[26] Lee Xiong, Chenyuan Xiong, Ye Li, Kowok-Fung Tang, Jialin Liu, Paul N. Bennett, Juna Ahmed, and Arnold Overwijk. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In 9th International Conference on Learning Representations. OpenReview.net.

[27] Jinxi Xu and W. Bruce Croft. 1996. Query Expansion Using Local and Global Document Analysis. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 4–11.

[28] HongChien Yu, Zhiyuan Dai, and Jamie Callan. 2021. PGT: Pseudo Relevance Feedback Using a Graph-Based Transformer. In Advances in Information Retrieval - 43rd European Conference on IR Research (Lecture Notes in Computer Science, Vol. 12657) Springer, 3887–3896.

[29] Xiaoyang Chen, Kai Hui, Ben He, Xianpei Han, Le Sun, and Zheng Ye. 2021. Co-BERT: A Context-Aware BERT Retrieval Model Incorporating Local and Query-specific Context. arXiv preprint arXiv:2104.08323 (2021).

[30] 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 30: Annual Conference on Neural Information Processing Systems. 5998–6008.

[31] Xiao Wang, Craig Macdonald, Nicola Tomellotito, and Iadh Ounis. 2021. Pseudo-Relevance Feedback for Multiple Representation Dense Retrieval. arXiv preprint arXiv:2106.11251.

[32] Chenyan Xiong, Zhiyuan 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, 55–64.

[33] Lee Xiong, Chenyuan Xiong, Ye Li, Kowok-Fung Tang, Jialin Liu, Paul N. Bennett, Juna Ahmed, and Arnold Overwijk. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In 9th International Conference on Learning Representations. OpenReview.net.

[34] Jinxi Xu and W. Bruce Croft. 1996. Query Expansion Using Local and Global Document Analysis. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 4–11.

[35] HongChien Yu, Zhiyuan Dai, and Jamie Callan. 2021. PGT: Pseudo Relevance Feedback Using a Graph-Based Transformer. In Advances in Information Retrieval - 43rd European Conference on IR Research (Lecture Notes in Computer Science, Vol. 12657) Springer, 3887–3896.