Domain Adaptation for Dense Retrieval and Conversational Dense Retrieval through Self-Supervision by Meticulous Pseudo-Relevance Labeling

Minghan Li and Eric Gaussier
mhli@suda.edu.cn, eric.gaussier@imag.fr

2: Domain Adaptation for Dense Retrieval

General problem formulation

$$RSV(q,d)_{DR} = g(q) \cdot g(d)$$

$$\text{or} \quad RSV(q,d)_{DR} = cos\big(g(q), g(d)\big)$$

Limitation Recent studies like BEIR [1] showed that dense retrieval [2] models trained on a source domain generalize less well than traditional models as BM25 and interaction-based models on out-of-distribution (OOD) data sets.

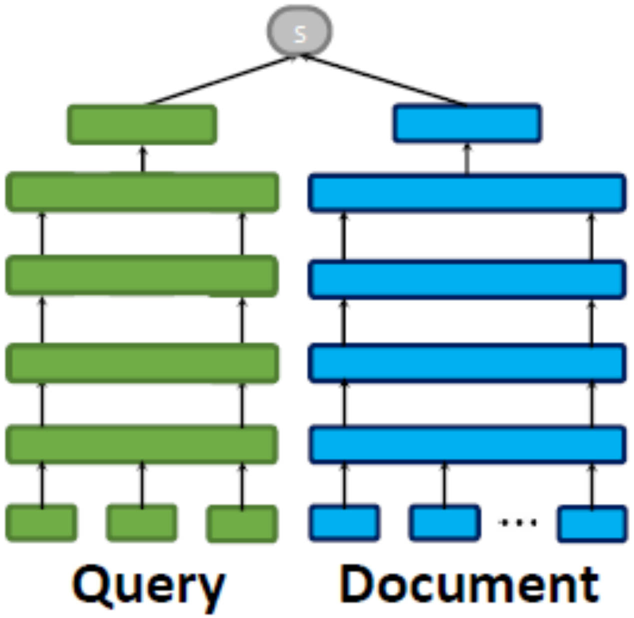


Figure: dense retrieval, figure from [3]

[1] Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. "BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models". In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2). 2021

[2] Karpukhin, Vladimir, et al. "Dense Passage Retrieval for Open-Domain Question Answering." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.

[3] Khattab, O., & Zaharia, M. (2020, July). Colbert: Efficient and effective passage search via contextualized late interaction over bert. In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval (pp. 39-48).

Université Grenoble Alpes

Related Work

- **Query generation approaches** [3,4,5]: a large sequence-to-sequence (seq2seq) model is used to generate queries for target domain data. This seq2seq model is trained on source domain.
- Alternative strategies: **MoDIR** [6] train a domain classifier that distinguishes source and target domains. The dense retrieval encoder is then trained in an adversarial manner to learn domain-invariant representations.
  - The results vary from one data set to the other, with sometimes important improvements and sometimes marginal gains or losses.

[3] Ji Ma, Ivan Korotkov, Yinfei Yang, Keith Hall, and Ryan McDonald. "Zero-shot Neural Passage Retrieval via Domain-targeted Synthetic Question Generation". In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021, pp. 1075–1088
[4] Davis Liang, Peng Xu, Siamak Shakeri, et al. "Embedding-based zero-shot retrieval through query generation". In: arXiv preprint arXiv:2009.10270 (2020)
[5] Kexin Wang, Nandan Thakur, Nils Reimers, and Iryna Gurevych. "GPL: Generative Pseudo Labeling for Unsupervised Domain Adaptation of Dense Retrieval". In NAACL2022, July 2022, pp. 2345–2360
[6] Xin, Ji, et al. "Zero-shot dense retrieval with momentum adversarial domain invariant representations." arXiv preprint arXiv:2110.07581 (2021).

Pseudo-Relevance Labeling for Dense Retrieval

Pseudo-Positive Sampling

- We simply propose here to consider, for each query, the top k documents obtained with the combination BM25&T53B, in which T53B [7] serves as a re-ranker, as relevant (or positive).
- T53B, which has been shown to be a good zero-shot IR model in [8], is fine-tuned on MS MARCO collection.

T53B

- [7] proposed to use T5 as an interaction-based model for information retrieval by relying on the following input representation:

Query: [q] Document: [d] Relevant: true or false

- The relevance score for inference is then determined by the likelihood of producing "true":

$$RSV(q,d)_{T5} = softmax(Z_{true}) = \frac{e^{Z_{true}}}{e^{Z_{true}} + e^{Z_{false}}},$$

[7] Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." The Journal of Machine Learning Research 21.1 (2020): 5485-5551.
[8] Nogueira, Rodrigo, Zhiying Jiang, and Jimmy Lin. "Document ranking with a pretrained sequence-to-sequence model." arXiv preprint arXiv:2003.06713 (2020).

What We Have Now?

All documents

- **in the dataset**
- **Global negative sampling**

BM25 list

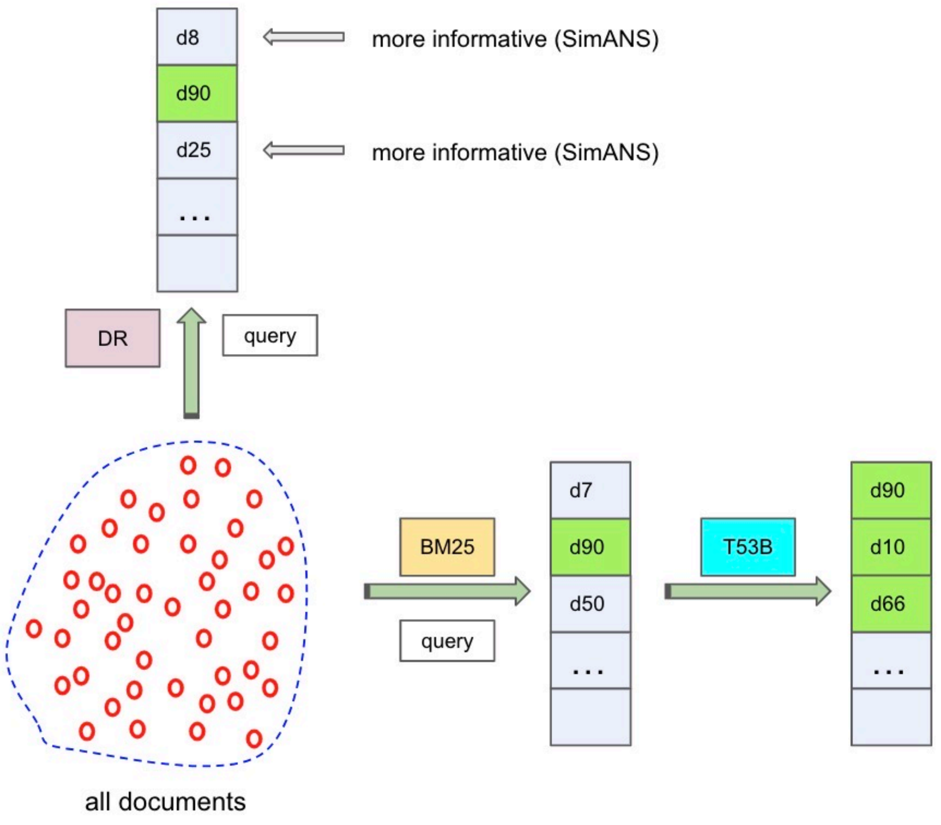
- **Fast Recall the relevant docs in the dataset**
- **Bm25 hard negative sampling**

T53B Model, reranking list

- **A reranker model, can rerank BM25 top N list**
- **Pseudo-positive sampling**

Dense Retrieval Model

- **SimANS hard negative sampling**



Global and BM25 Hard Negative Sampling

Generating Positive-Negative Training Pairs on the Target Domain

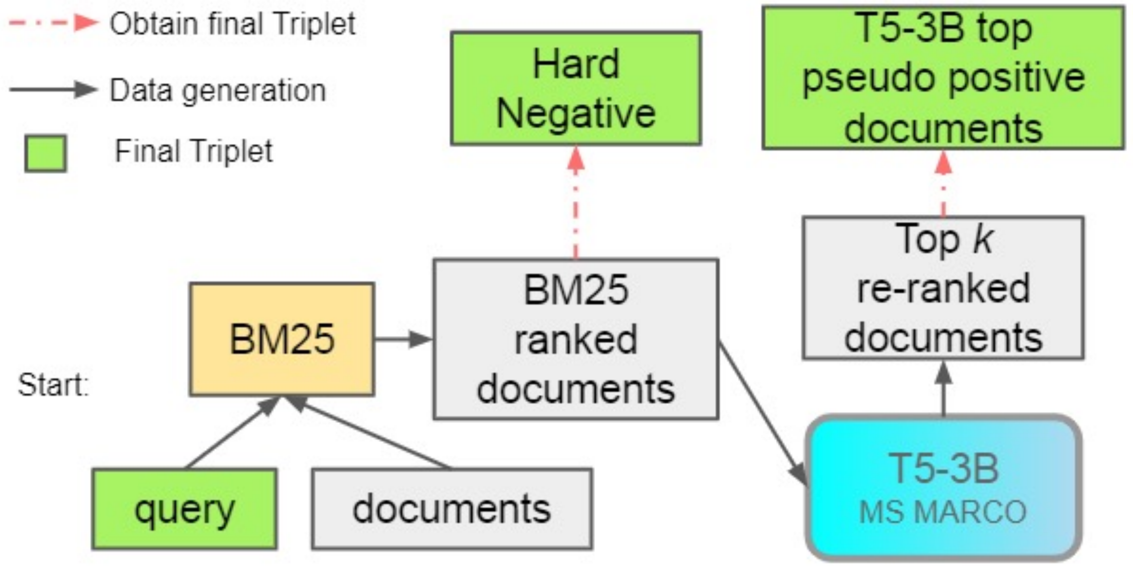


Fig. 4.1.: The overall pipeline of generating self-supervised data with BM25 hard negative sampling for pseudo-relevance labeling.

Global Negative vs BM25 Hard Negative

SimANS Hard Negative Sampling

Generating Positive-Negative Training Pairs on the Target Domain

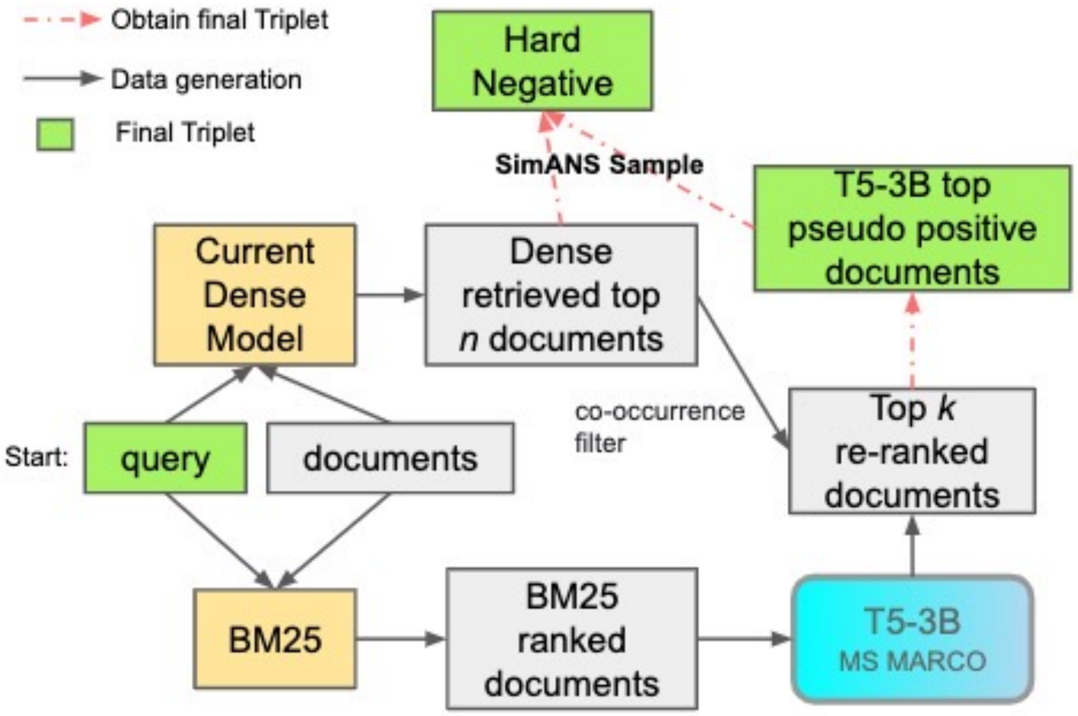


Figure 2: The overall pipeline of generating self-supervised data with meticulous pseudo-relevance labeling using SimANS hard negative sampling.

$$p_i \propto \exp\left(-a\left(s(q, d_i) - s(q, \tilde{d}^+) - b\right)^2\right), \forall\, d_i \in \widetilde{\mathcal{D}}^-, \quad (1)$$

Kun Zhou, Yeyun Gong, Xiao Liu, et al. "SimANS: Simple Ambiguous Negatives Sampling for Dense Text Retrieval". In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track. Abu Dhabi, UAE: Association for Computational Linguistics, Dec. 2022, pp. 548–55

Results

Table 2: Domain adaptation result of FiQA, BioASQ and Robust04 (during training only use train queries).

Method	FiQA	BioASQ	Robust04	Avg.
Zero-Shot Models				
D-BERT	26.7	53.6	39.1	39.8
BM25 (Anserini)	23.6	73.0	44.4	47.0
Re-Ranking with Cross-Encoders (Upper Bound)				
BM25 + CE	33.1	72.8	45.8	50.6
BM25 + T53B	39.2	76.1	51.8	55.7
Previous Domain Adaptation Methods				
UDALM	23.3	33.1	-	-
MoDIR (ANCE)	29.6	47.9	-	-
Pre-Training based: Target $\rightarrow$ D-BERT				
SimCSE	26.7	53.2	-	-
ICT	27.0	55.3	-	-
TSDAE	29.3	55.5	-	-
Generation-based (Previous SOTA)				
QGen	28.7	56.5	-	-
GPL	32.8	62.8	41.9	45.8
TSDAE + GPL	34.4	61.6	40.7	45.6
Proposed: T53B, Global Random Neg				
DoDress-T53B (D-BERT)	27.3	52.9	40.5	40.2
DoDress-T53B (GPL)	33.0	62.0	43.2	46.1
Proposed: T53B, BM25 Hard Neg				
DoDress-T53B (D-BERT)	30.4	58.6	41.6	43.5
DoDress-T53B (GPL)	34.2	64.7	43.3	47.4
Proposed: T53B, SimANS Hard Neg				
DoDress-T53B (D-BERT)	**31.0**	**60.6**	**43.6**	**45.1**
DoDress-T53B (GPL)	**34.9**	**65.3**	**45.5**	**48.6**

2: Domain Adaptation for Conversational Search

Conversational document search, which is to find relevant documents from collections of documents in response to user queries in a conversational context, is often referred to as "conversational search" as documents are the typical output generated by the system.

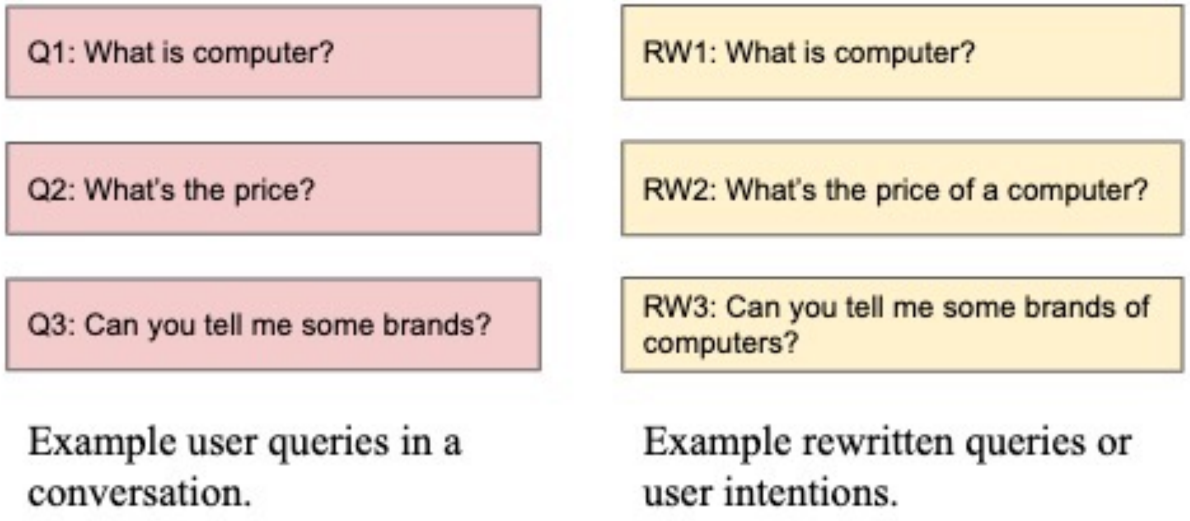


Figure 4: An example of conversational search queries.

Challenges

- Conversations exhibit contextualization, conciseness, and reliance on prior knowledge, presenting challenges for search systems in accurately understanding information needs
- More ambiguous, often containing references and omissions from previous turns
- Data scarcity, especially DR models require large amount of data

Université Grenoble Alpes

Related Work

ConvDR

- Learning from an well trained ad hoc dense retriever as teacher, to mimic the teacher embeddings on oracle reformulated queries on CANARD

CQE

- Use the conversational queries and human rewritten queries in the CANARD for the target datasets (documents)

While these approaches still face domain gaps in the training data.

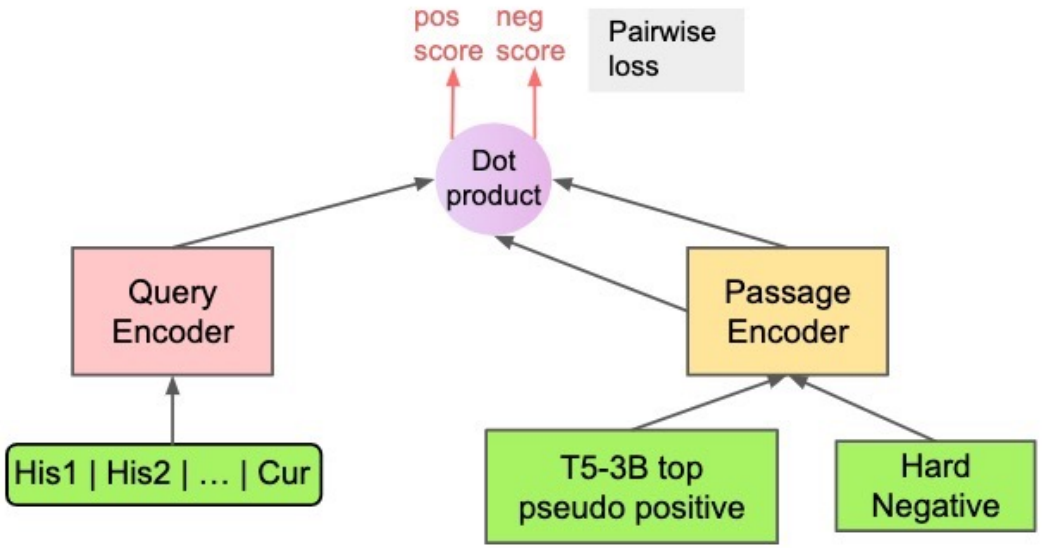


Fig. 5.3.: After generating pseudo-labeling data, now do domain adaptation for the dense retrieval model for target conversational search corpus. Now the encoders of the dense model (pre-trained on MS MARCO) are no longer shared, and the query encoder is trained to generate whole representation of the conversation (concatenated, not rewritten by T5-Large, since for fast online search).

Pseudo-Relevance Labeling for Conversational Dense Retrieval

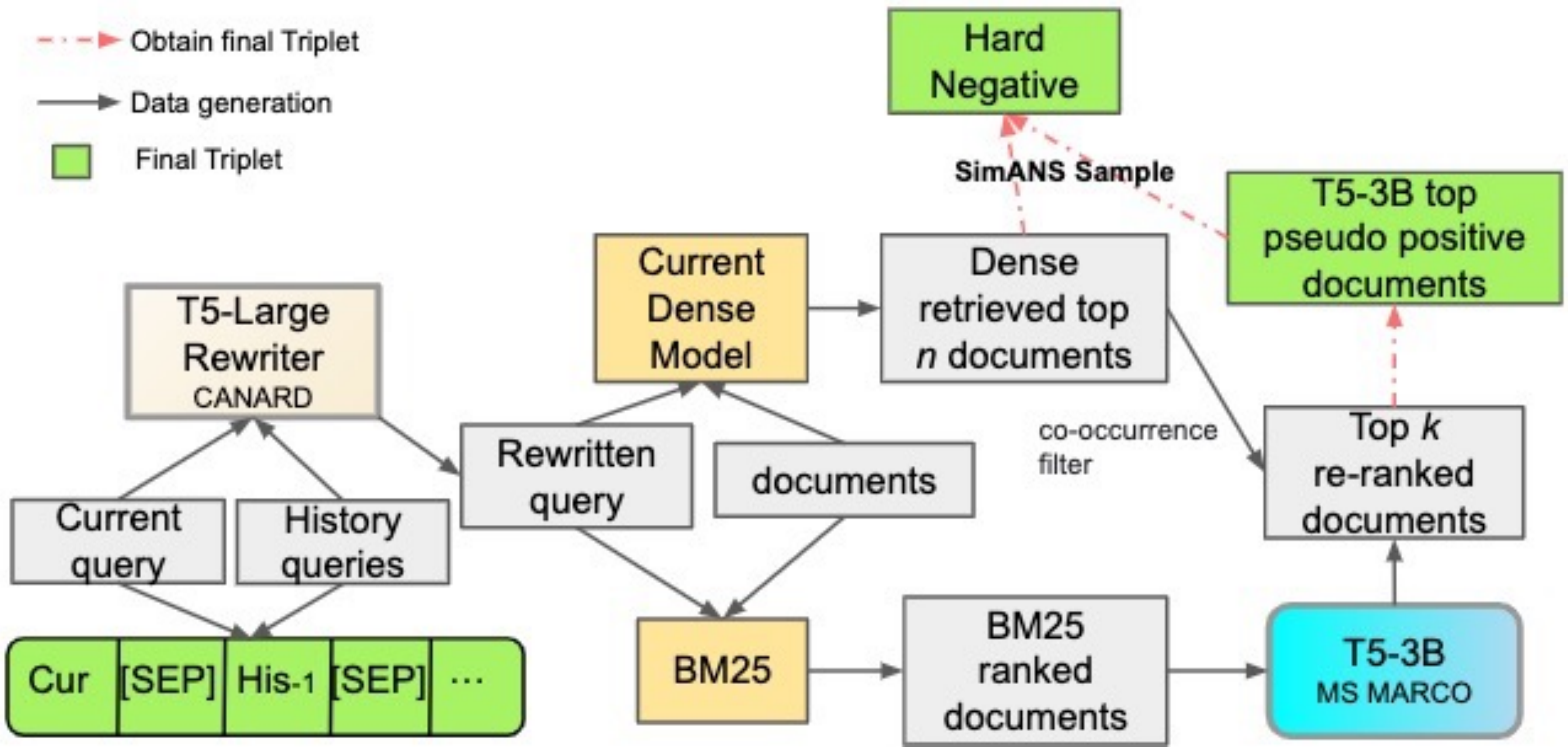


Figure: The T5-Large rewrites the conversational query to a human language style sentence. Then for each turn, the rewritten query is used for generating pseudo-labels, where the steps are similar as dense retrieval's.

Result

Table 4: Domain adaptation result of CAsT-19.

model	nDCG@3 (%)
Zero-Shot Models	
BERT-dot-v5(current)	33.4
BERT-dot-v5(concatenation)	27.2
BERT-dot-v5(T5Rewrite)	53.2
BERT-dot-v5(Human) (Upper Bound)	58.9
BM25(Human)	37.0
BM25(T5Rewrite)	31.2
Re-Ranking with Cross-Encoders	
T5-3B rerank T5Rewrite	56.7
Related Work	
ConvDR (BERT-dot-v5)	55.4
CQE (BERT-dot-v5)	53.7
Proposed Approach	
T53B, SimANS Neg, based on ConvDR	
DoDress-T53B (BERT-dot-v5)	**58.0**
T53B, SimANS Neg, based on CQE	
DoDress-T53B (BERT-dot-v5)	**57.6**

Thanks!