Algorithm 1 Fine-Tuning Method for Document-Level Passage Retrieval

Require:

Input:

query_encoder: Transformer-based encoder for queries

passage_encoder: Transformer-based encoder for passages

D: Single document from which query-passage pairs are to be extracted

t: Learned temperature parameter

n: Batch size for training

d: Embedding dimension

Ensure:

Output:

Loss: Computed symmetric loss

procedure DOCUMENT_LEVEL_PASSAGE_RETRIEVAL(D, t, n, d)

  $Q, P \leftarrow D.get\_pairs(num\_pairs = n, unique\_passages = True)$

  $Q_e \leftarrow query\_encoder(Q)$ ▷ Dimension: $n \times d$

  $P_e \leftarrow passage\_encoder(P)$ ▷ Dimension: $n \times d$

  $logits \leftarrow dot\_product(Q_e, P_e^\top) \times \exp(t)$ ▷ Dimension: $n \times n$

  $labels \leftarrow Sequence from 1 to n$

  $loss\_q \leftarrow cross\_entropy(logits, labels)$

  $loss\_p \leftarrow cross\_entropy(logits^\top, labels)$

  $Loss \leftarrow \frac{loss\_q + loss\_p}{2}$

end procedure