



Jihyuk Kim



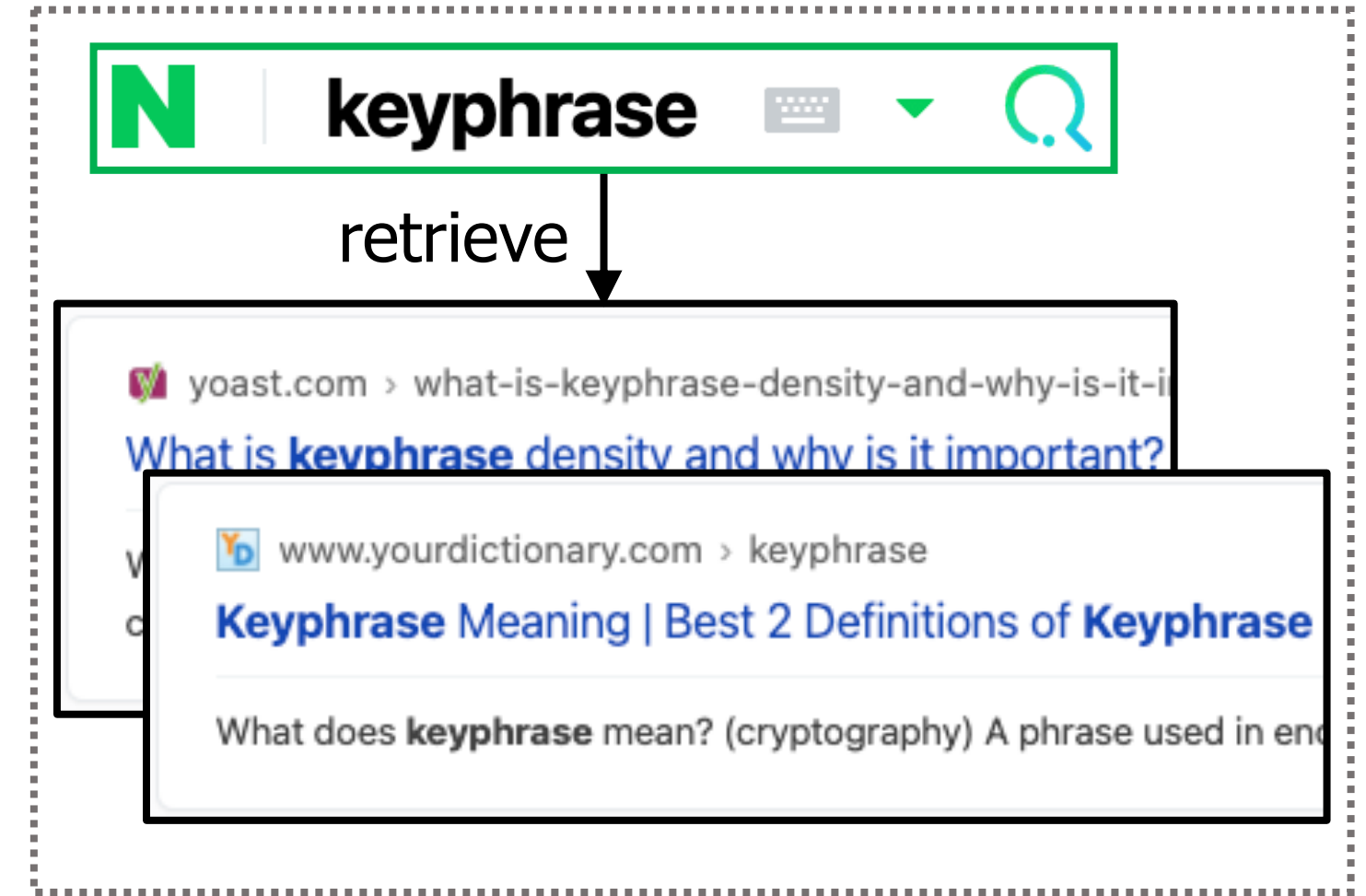
Minsoo Kim

Seung-won Hwang



Summary

- ❖ We study **passage retrieval task** using MSMARCO and TREC-DL'19/20 datasets.
- ❖ We tackle the challenge of **incomplete relevance annotation**.
- ❖ We adopt **knowledge distillation**, and introduce **collective bi-encoder teacher**.



Task: Information Retrieval (IR)

IR aims to retrieve relevant passages to the given query.
IR systems can be used not only for search engines but also for retrieval-augmented approaches.

Challenge: Incomplete Relevance Annotation

Q. Do we have enough labels?

A. No, we don't. Since there are too many passages (e.g., 8.8M) to be annotated, it is challenging to annotate all relevant passages.

Dataset	# of passages	# of queries	# of relevant passages per query
MSMARCO Train set	8.8M	500k	1.1 → Only tiny portions are annotated!
TREC-DL 2019 Test set		43	58.2

+ What if new relevant passages come in without annotation?

Approach: Knowledge Distillation (KD)

KD introduces high-capacity teachers, to provide pseudo-relevance labels.



Teacher		Capacity	Efficiency
Baseline	cross-encoder	✓	✗
	bi-encoder	✗	✓
Ours	collective bi-	✓	✓

Q/P	Content	Relevance Prediction	
		Individual	Collective
Q	Is caffeine an analgesic?		
Top-1 P ✓	... caffeine , and codeine ... treat headaches.	High	High
Top-2 P ✓	butalbital, and caffeine ... treat tension headaches.	High	High
Top-3 P ✗	... Caffeine is considered a safe ingredient. ...	Low	Low
Collective knowledge	<i>treat tension headaches</i>		

Experiment: Passage Retrieval

Dataset

- We use public benchmark datasets: MSMARCO and TREC-DL test collections.

Comparison

- w/o KD (-)
- KD from a cross-encoder (**CE**)
- KD from an ensemble of CE (**CE Ensemble**)
- KD from collective bi-encoder (**Ours**)

Evaluation Results

- 1) **Teachers** provide more informed labels than incomplete labels.
- 2) **Our collective bi-encoder teacher** is more effective than cross-encoder teachers.

