Phrase Retrieval Learns Passage Retrieval, Too

Jinhyuk Lee[1,2]   **Alexander Wettig**[1]   **Danqi Chen**[1]
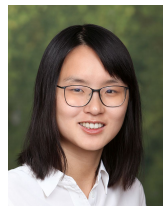
[1] Princeton University
[2] Korea University

Open-Domain Question Answering

Question
How many Warsaw's inhabitants spoke Polish in 1933?

Document Retriever

Warsaw

Document Reader

833,500

Retriever-Reader
(Chen et al., 2017)

100 passages with
BERT-base (110M)

DPR (Karpukhin et al., 2020)

⋮

100 passages with
T5-large (770M)

FiD (Izacard and Grave, 2020)

Phrase Retrieval for Open-Domain QA
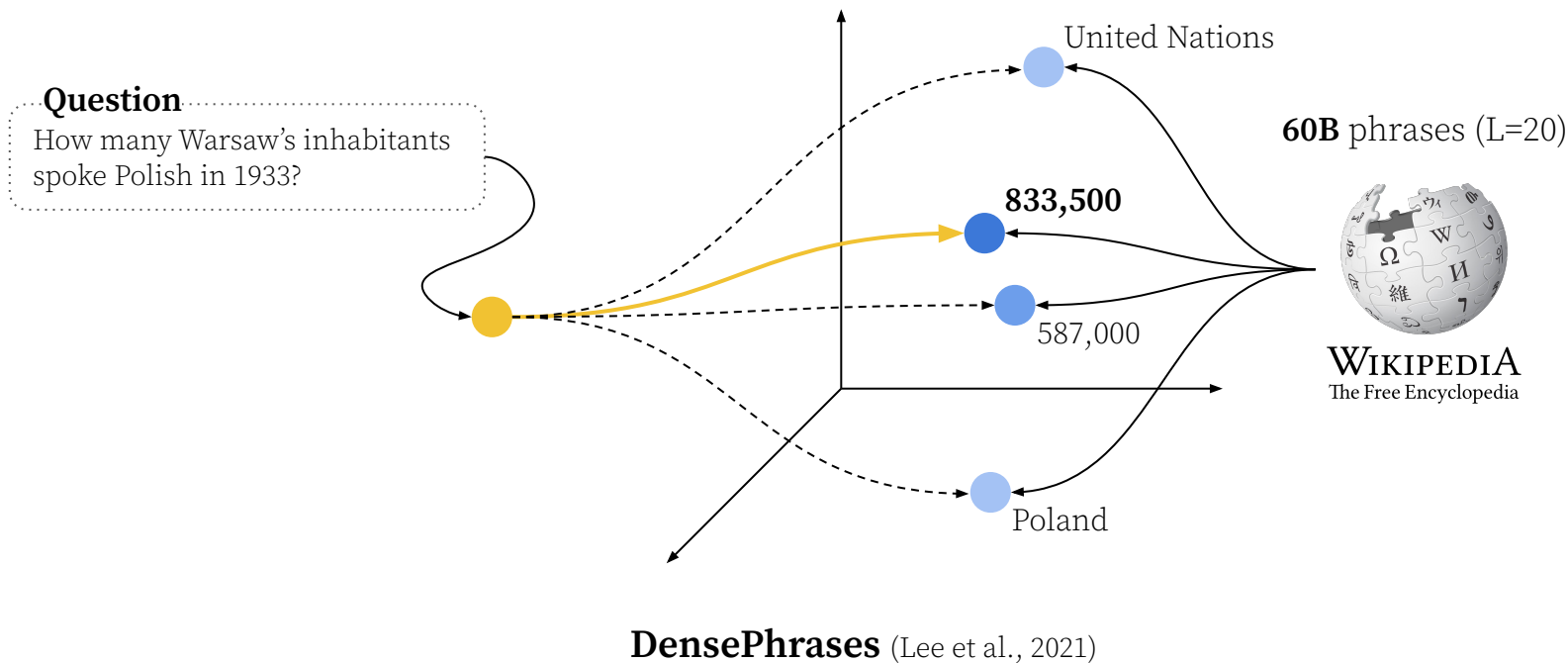
Phrase = any contiguous segment of text up to L words (Seo et al., 2019)

Question

How many Warsaw's inhabitants spoke Polish in 1933?

United Nations

60B phrases (L=20)

833,500

587,000

Poland

WIKIPEDIA
The Free Encyclopedia

DensePhrases (Lee et al., 2021)
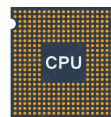
Phrase Retrieval is **Accurate** and **Fast**
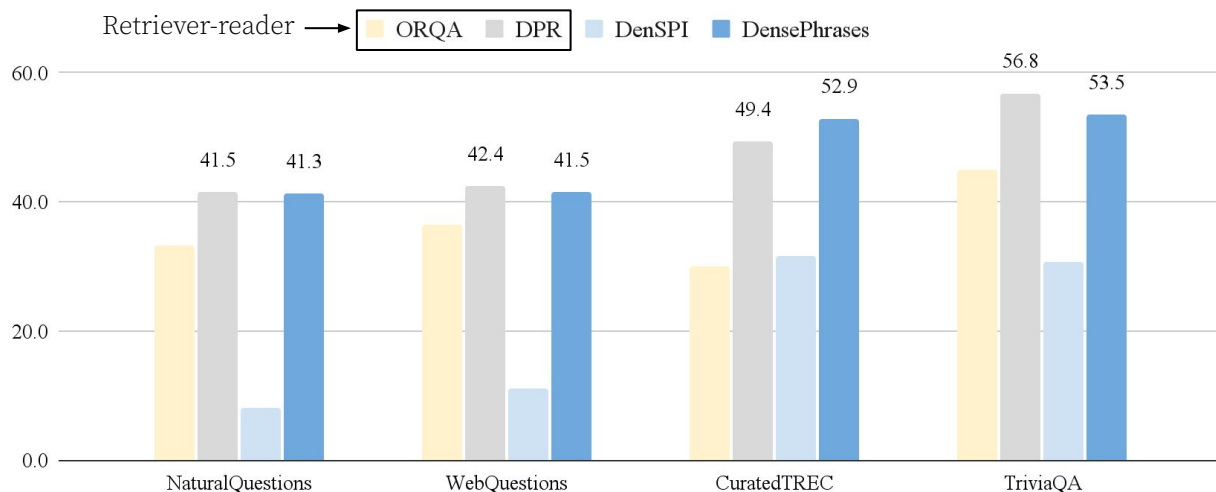
Without any reader model, phrase retrieval is **competitive** with retriever-reader approaches.

0.04 Q/sec
(DPR)

<

13.6 Q/sec
(DensePhrases)

Dense phrase retrieval makes open-domain QA **fast** and **simple**!

Fixed Granularity for Text Retrieval
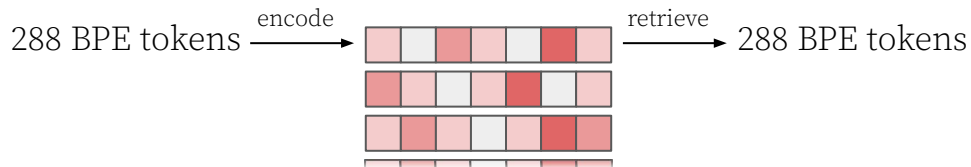
Sentence Retrieval

SBERT (Reimers et a., 2019)**,** SimCSE (Gao et al., 2021): **1 sentence**

Passage Retrieval

ORQA (Lee et al., 2019): **288 BPE tokens** for a passage

DPR (Karpukhin et al., 2020): **100 words** for a passage

288 BPE tokens $\xrightarrow{\text{encode}}$ $\xrightarrow{\text{retrieve}}$ 288 BPE tokens

Different index for **different** granularity?

Phrases as a Basic Retrieval Unit

| Examples ▾ | Why is the sky blue? | 🔍 |

The blue color is sometimes wrongly attributed to **Rayleigh scattering,** which is responsible for the color of the sky. Rather, water ice is blue for the same reason that large quantities of liquid water are blue: it is a result of an overtone of an oxygen–hydrogen (O−H) bond stretch in water, which absorbs light at the red end of the visible spectrum. In the case of oceans or lakes, some of the light hitting the surface of water is reflected back directly, but most of it penetrates the surface, interacting with its molecules. The water molecule can vibrate in different modes when light hits it. The red, orange, yellow, and green wavelengths of light are absorbed so that the remaining light is composed of the shorter wavelengths of blue and violet. This is the main reason why the ocean is blue. So, water owes its intrinsic blueness to selective absorption in the red part of its visible spectrum. The absorbed photons promote transitions to high overtone and combination states of the nuclear motions of the molecule, i.e. to highly excited vibrations. An example of blue ice was observed in Tasman Glacier, New Zealand in January 2011. Blue ice is exposed in areas of the Antarctic where there is no net addition or subtraction of snow. That is, any snow that falls in that area is counteracted by sublimation or other losses. These areas have been used as runways (e.g. Wilkins Runway, Novolazarevskaya, Patriot Hills Base Camp) due to their hard surface, which is suitable for aircraft fitted with wheels rather than skis.

↑
Document Title

Retrieving **Phrases** ⇒ Sentences ⇒ Passages ⇒ Documents ⇒ ...
⇍ ⇍ ⇍ ⇍

Single index for **multi** granularity!

In This Paper …
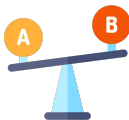
Q1: Is this **better** than passage retrievers?
Experiment #1: Passage Retrieval / Experiment #2: Open-domain QA

Q2: **Why** does this work?
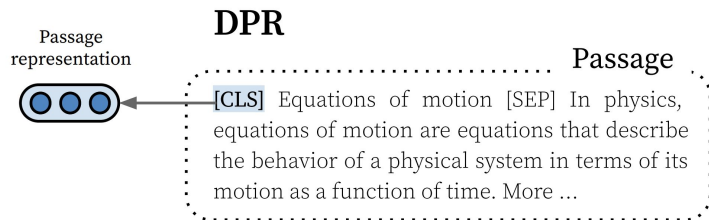Analysis / Experiment #3: Entity Linking & Dialogue

Q3: How **efficient** is this?
Phrase Filtering & Quantization-aware Fine-tuning

Formulation

Passage Retrieval

DPR

Passage representation

Passage

[CLS] Equations of motion [SEP] In physics, equations of motion are equations that describe the behavior of a physical system in terms of its motion as a function of time. More …

Question vector

$$f(p, q) = E_p(p)^\top E_q(q)$$

Single vector for each passage

Phrase-based Passage Retrieval

DensePhrases

Passage representation

Passage

[CLS] Equations of motion [SEP] In physics, equations of motion are equations that describe the behavior of a physical system in terms of its motion as a function of time. More …
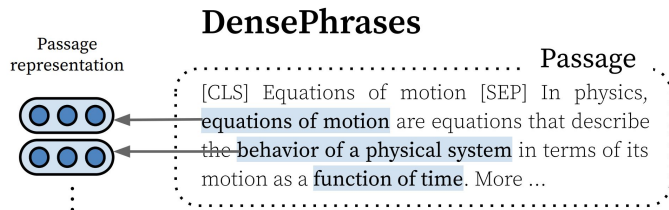
Phrase vector

$$\tilde{f}(p, q) := \max_{s^{(p)} \in \mathcal{S}(p)} E_s(s^{(p)})^\top E_q(q)$$

Multiple (phrase) vectors for each passage

Passage Retrieval: DPR vs DensePhrases
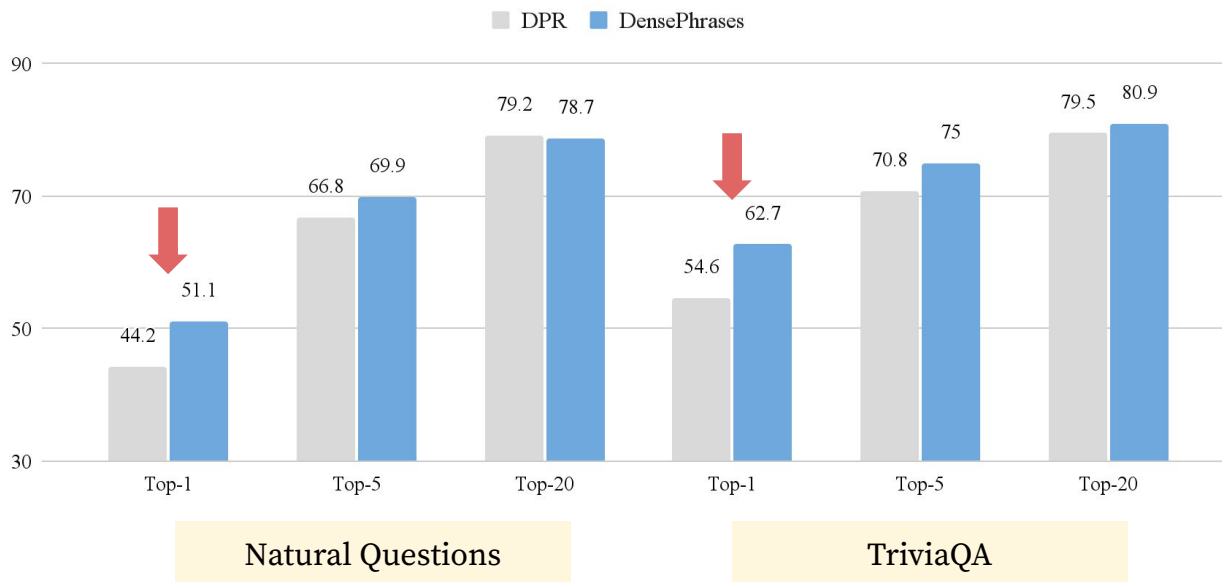

Without any re-training, **DensePhrases outperforms DPR** on passage retrieval!

Larger gains when **k is small**.

Fusion-in-Decoder for Open-domain QA

Izacard and Grave, 2021
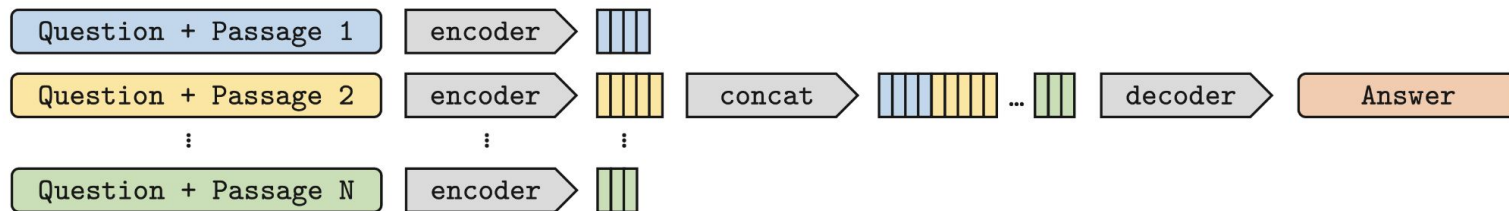
Feeds top-k passages from **DPR** to **T5** (Raffel et al., 2020) to generate answers.

FiD achieves state-of-the-art performance **when k is large** (e.g., k=100).

Requires 64 **32GB** V100 GPUs for training!

Feed top-k passages from **DensePhrases** to T5 to generate answers?

Open-domain QA: DPR vs DensePhrases

DensePhrases outperforms DPR on open-domain QA (+6.4 EM when k=5).
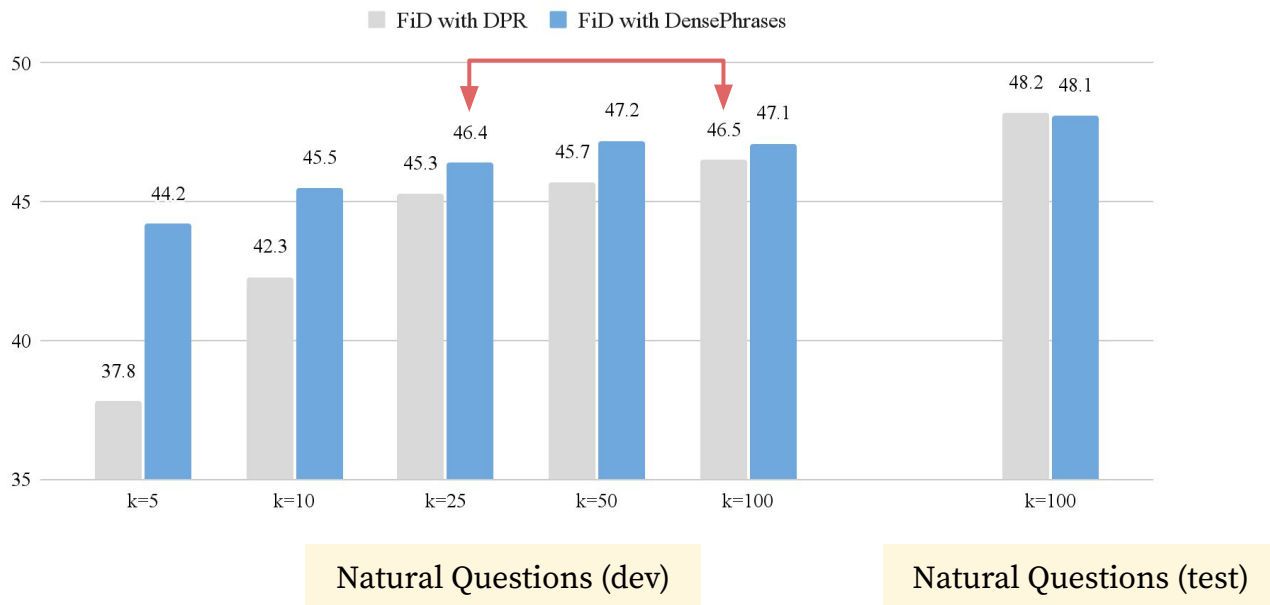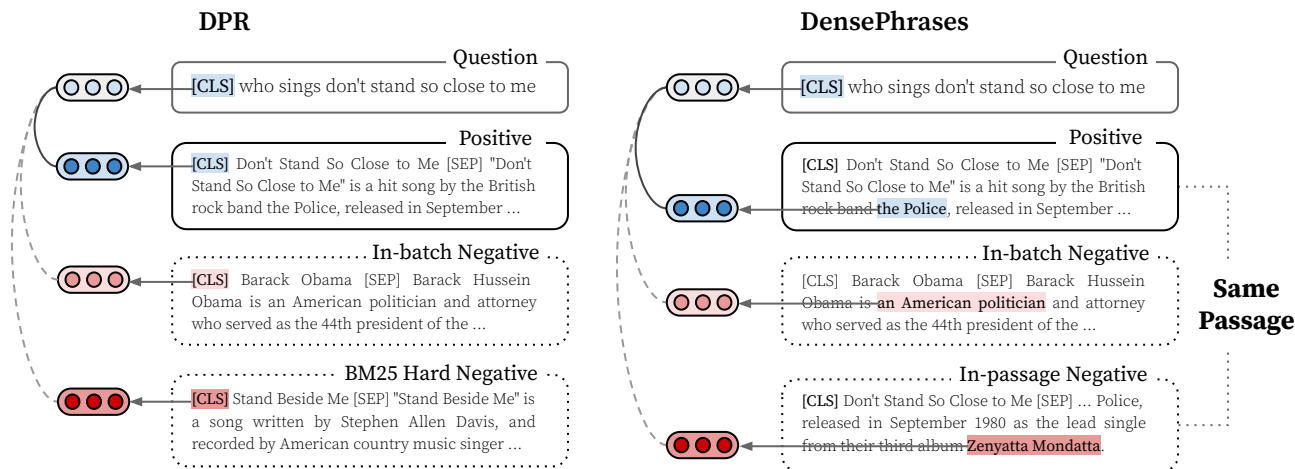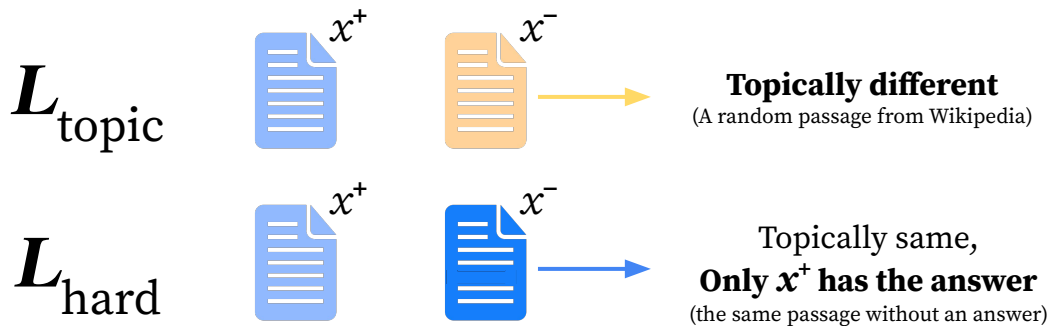
k=25~50 is enough for good performance (k ≤ 50 fits in **24GB**)

Why DensePhrases > DPR on Passage Retrieval?

DPR

DensePhrases

Question

[CLS] who sings don't stand so close to me

Question

[CLS] who sings don't stand so close to me

Positive

[CLS] Don't Stand So Close to Me [SEP] "Don't Stand So Close to Me" is a hit song by the British rock band the Police, released in September ...

Positive

[CLS] Don't Stand So Close to Me [SEP] "Don't Stand So Close to Me" is a hit song by the British rock band the Police, released in September ...

In-batch Negative

[CLS] Barack Obama [SEP] Barack Hussein Obama is an American politician and attorney who served as the 44th president of the ...

In-batch Negative

[CLS] Barack Obama [SEP] Barack Hussein Obama is an American politician and attorney who served as the 44th president of the ...

Same Passage

BM25 Hard Negative

[CLS] Stand Beside Me [SEP] "Stand Beside Me" is a song written by Stephen Allen Davis, and recorded by American country music singer ...

In-passage Negative

[CLS] Don't Stand So Close to Me [SEP] ... Police, released in September 1980 as the lead single from their third album Zenyatta Mondatta.

In-passage negatives in DensePhrases work similar to
BM25 hard negatives in DPR!

Analysis with L_{topic} and L_{hard}

L_{topic} x^+ x^- **Topically different**
(A random passage from Wikipedia)

L_{hard} x^+ x^- Topically same,
Only x^+ has the answer
(the same passage without an answer)

For both metrics, lower numbers are better.

DPR has good L_{topic} while **DensePhrases** has good L_{hard}.

L_{topic} and L_{hard} : What Do They Really Mean?

DPR (Karpukhin et al., 2020)

Where is Princeton University located?		Run

Title: *Princeton University* ⟶ topically relevant!   Retrieval ranking: #2 $P(p|q)$=0.43 $P(a|p,q)$=0.94 $P(a,p|q)$=0.41 ✅

... "Cherokee Advocate", graduated in 1844. Princeton University Princeton University is a private Ivy League research university in **Princeton, New Jersey**. Founded in

http://qa.cs.washington.edu:2020/

DensePhrases (Lee et al., 2021)

Examples ▾   Where is Princeton University located?   🔍

19 results (106ms)   ☑ Real-time Search   English Wikipedia (2018.12.20)

The New York metropolitan area is home to many prestigious institutions of higher education. Three Ivy ✅   [New York metropolitan area]
League universities: Columbia University in Manhattan, New York City; Princeton University in **Princeton,**
New Jersey; Yale University in New Haven, Connecticut – all ranked amongst the top 3 U.S. national

http://densephrases.korea.ac.kr

Good L_{hard} can give correct answer
even when **the passage is less relevant**.

topically less relevant,
but still correct answer!

14

But WHAT IF...? L_{topic} Matters?

For many coarse-granularity retrieval,
we need good L_{topic} !

Entity Linking

United Nations Security Council

[START_ENT] **Security Council** [END_ENT] members expressed concern on Thursday.
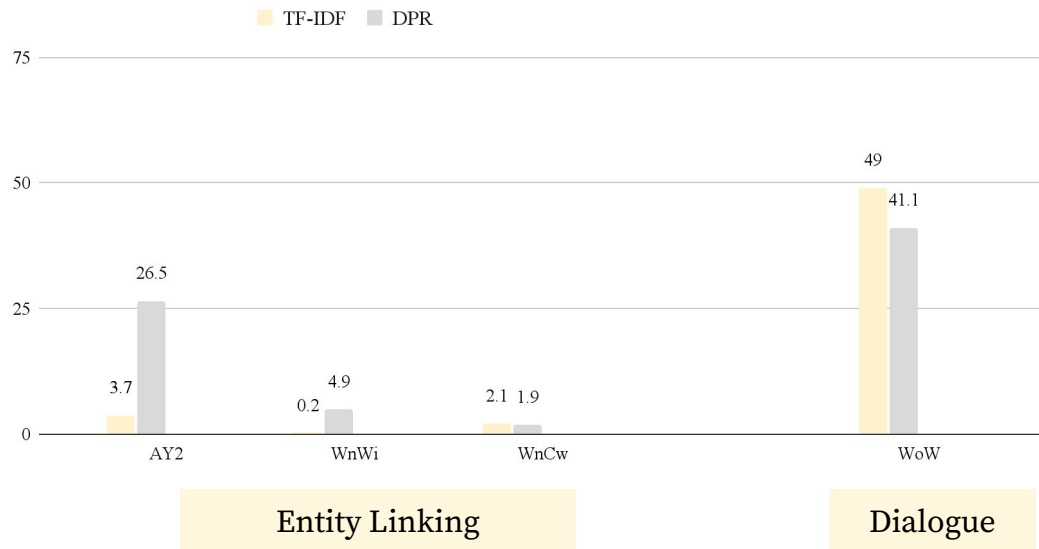
Knowledge-grounded Dialogue

Yamaha Corporation

Have you heard of Yamaha? They started as a piano manufacturer in 1887!

Only **one document** is relevant (annotated) for each query!
(KILT; Petroni et al., 2021)

Retrieval for **Entity Linking & Dialogue**

Retrieval for **Entity Linking & Dialogue**

Examples ▾ | do you like hip hop? | 🔍

20 results (93ms)   ☑ Real-time Search   English Wikipedia (2018.12.20)

Radio DJs or radio personalities introduce and play music that is broadcast on AM, FM, digital or Internet radio stations. Club DJs, commonly referred as DJs in general, play music at musical events, such as parties at music venues or bars, music festivals, corporate and private events. Typically, club DJs mix music recordings from two or more sources using different mixing techniques in order to produce non-stopping flow of music. One key technique used for seamlessly transitioning from one song to another is beatmatching. A DJ who mostly plays and mixes one specific music genre is often given the title of that genre; for example, a DJ who plays **hip hop music** is called a hip hop DJ, a DJ who plays house music is a house DJ, a DJ who plays techno is called a techno DJ, and so on. The quality of a DJ performance (often called a DJ mix or DJ set) consists of two main features: technical skills, or how well can DJ operate the equipment and produce smooth transitions between two or more recordings and a playlist, or ability of a DJ to select most suitable recordings also known as "reading the crowd".

Disc jockey

$f(s|D,q)=90.02$

Hip-hop music has reached the cultural corridors of the globe and has been absorbed and reinvented around the world. Hip hop music expanded beyond the US, often blending local styles with hip hop. Hip hop has globalized into many cultures worldwide, as evident through the emergence of numerous regional scenes. It has emerged globally as a movement based upon the main tenets of hip hop culture. The music and the art continue to embrace, even celebrate, its transnational dimensions while staying true to the local cultures to which it is rooted. Hip-hop's impact differs depending on each culture. Still, the one thing virtually all hip hop artists worldwide have in common is that they acknowledge their debt to those **African-American people in New York** who launched the global movement.

Hip hop music

✅

$f(s|D,q)=88.63$

topically relevant!
(annotated)

Hip-hop music has reached the cultural corridors of the globe and has been absorbed and reinvented around the world. Hip hop music expanded beyond the US, often blending local styles with hip hop. Hip hop has globalized into many cultures worldwide, as evident through the emergence of numerous regional scenes. It has emerged globally as a movement based upon the main tenets of hip hop culture. The music and the art continue to embrace, even celebrate, its transnational dimensions while staying true to the local cultures to which it is rooted. Hip-hop's impact differs depending on each culture. Still, the one thing virtually all hip hop artists worldwide have in common is that they acknowledge their debt to those African-American people in **New York who launched the global movement.**

Hip hop music

✅

$f(s|D,q)=88.23$

topically relevant!
(annotated)

75

50

25

0

Maximize the marginal probability of
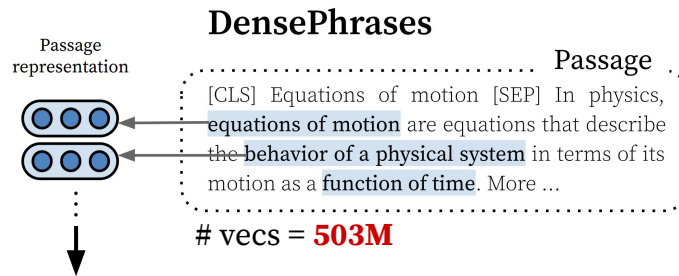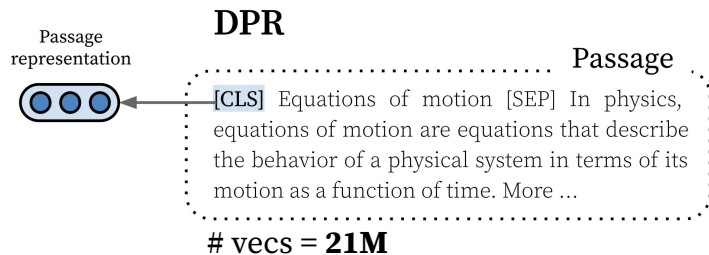**any phrases in the relevant document**

Retrieval for **Entity Linking & Dialogue**
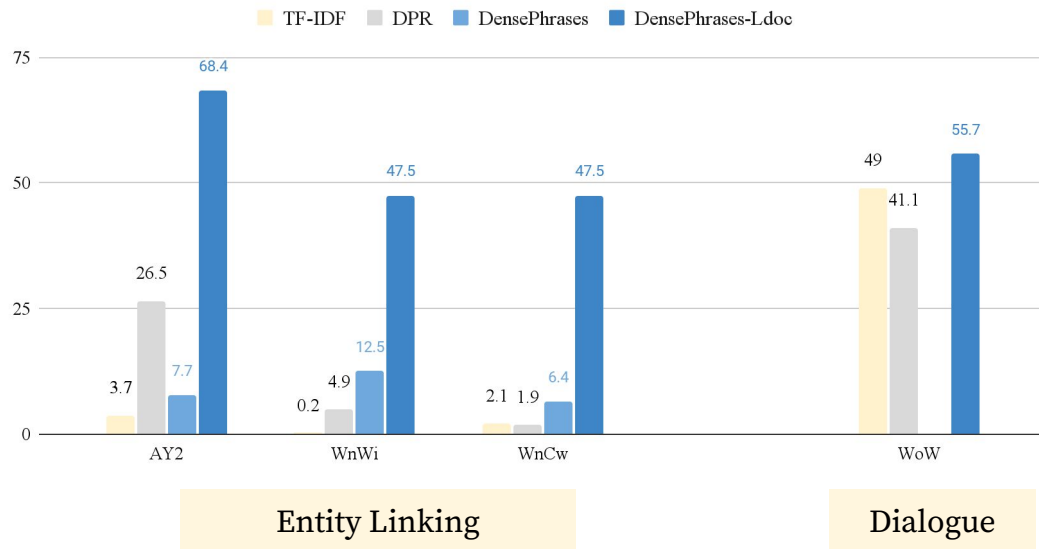
DensePhrases can be adapted to **retrieve topically relevant documents**!

Problem of **Multi-vector** Encoding

Luan et al., 2021; Khattab and Zaharia, 2020

DPR

Passage representation

[CLS] Equations of motion [SEP] In physics, equations of motion are equations that describe the behavior of a physical system in terms of its motion as a function of time. More ...

Passage

vecs = **21M**

DensePhrases

Passage representation

[CLS] Equations of motion [SEP] In physics, equations of motion are equations that describe the behavior of a physical system in terms of its motion as a function of time. More ...

Passage

vecs = **503M**

More vectors, **more space**!

Phrase indexes are **heavy**!

1.2TB (Seo et al., 2019)   **1.5TB** (Lee et al., 2020)   **320GB** (Lee et al., 2021)
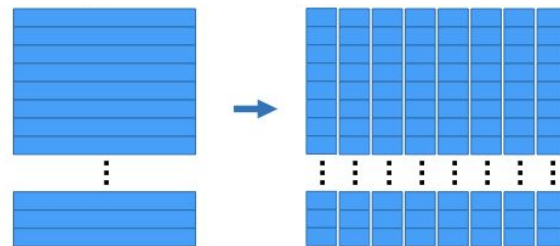
Reducing the Size of Phrase Index

"The New York metropolitan area is home to many prestigious institutions of higher education."

Phrase Filter
(Lee et al., 2021)

"**The New York metropolitan area**"

"prestigious institutions"

"higher education"

...

Optimized Product Quantization
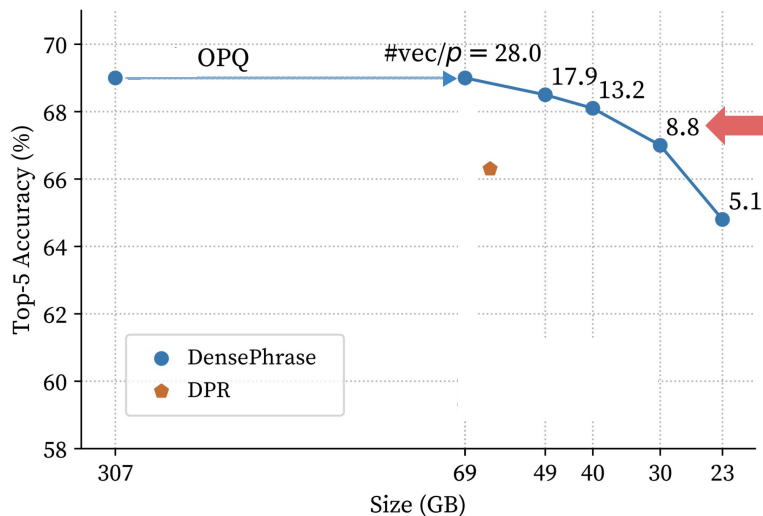(Ge et al., 2013)
+
Query-side Fine-tuning
(Lee et al., 2021)
=
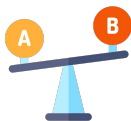**Quantization-aware Fine-tuning**

Reducing the Size of Phrase Index

We can safely reduce the size down to **23GB**! (DPR = 69GB)

DensePhrases with **# vector/passage = 8.8** is similar to DPR.

Conclusion

Q1: Is this **better** than passage retrievers?

Yes! **DensePhrases > DPR** on passage retrieval and open-domain QA!

Q2: **Why** does this work?

Better at **fine-grained entailment**, can be used for coarse retrieval.

Q3: How **efficient** is this?

Can safely reduce the index size from **307GB to 23GB**!

Paper: https://arxiv.org/abs/2109.08133
Code & Models: https://github.com/princeton-nlp/DensePhrases
Demo: http://densephrases.korea.ac.kr/
E-mail: jinhyuklee@cs.princeton.edu