



DSBA CS224n 2021 Study

# [Lecture 11]

## Question Answering

---



고려대학교 산업경영공학과  
Data Science & Business Analytics Lab

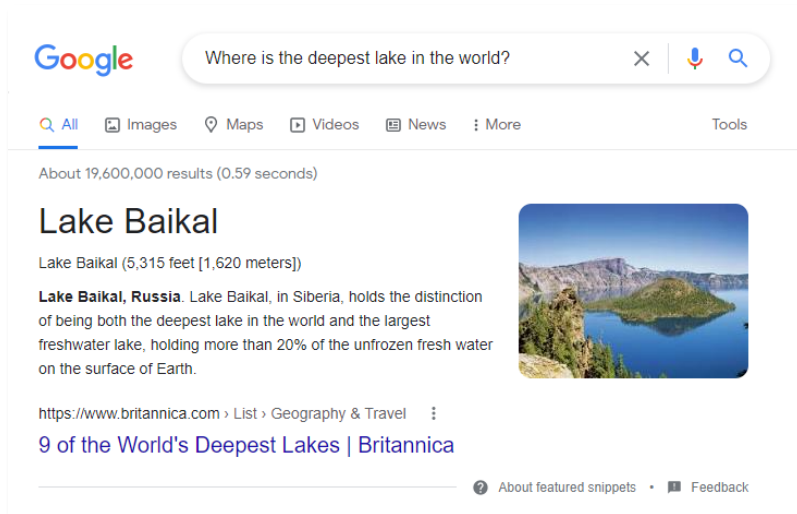
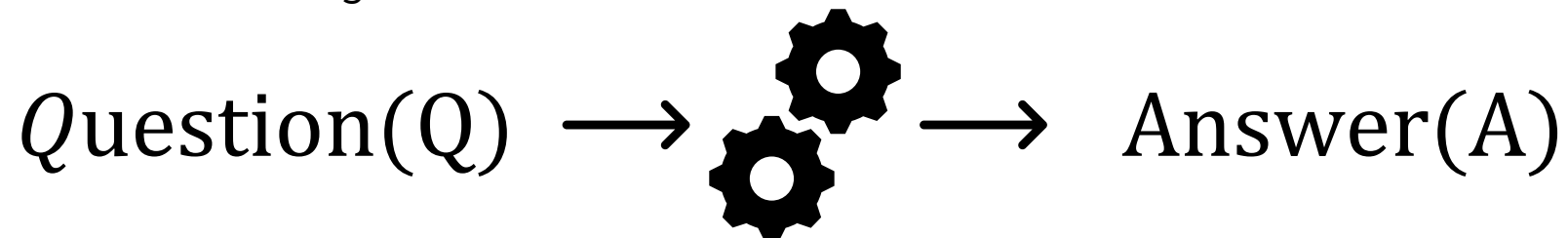
발표자 : 김선우

- 1 Question Answering
- 2 Reading comprehension
- 3 Open-domain (textual) question answering

## 01

# Question Answering

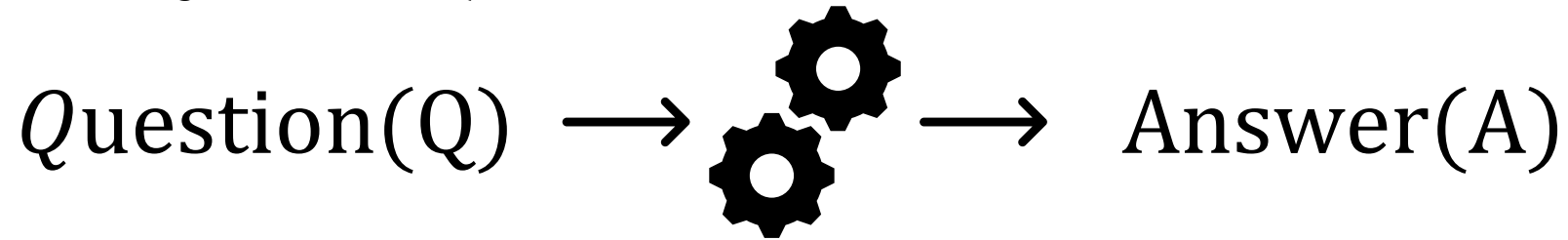
What is question answering?



사람의 언어로 된 질문에 자동적으로 답할 수 있는 시스템을 만들자!

# Question Answering

Question answering : a taxonomy



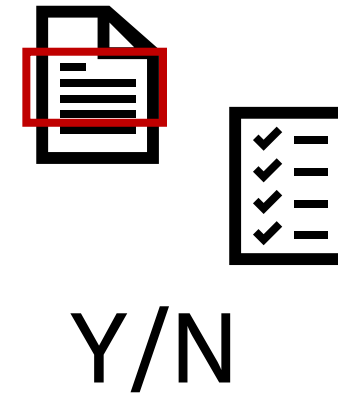
정보 소스



질문 유형

- ✓ Factoid vs non-factoid
- ✓ Open-domain vs closed domain
- ✓ Compositional vs simple

답 유형



# Question Answering

Question answering in deep learning era

대부분의 state-of-the-art question answering 시스템들은 end-to-end train과 pre-train된 language model 위해 build

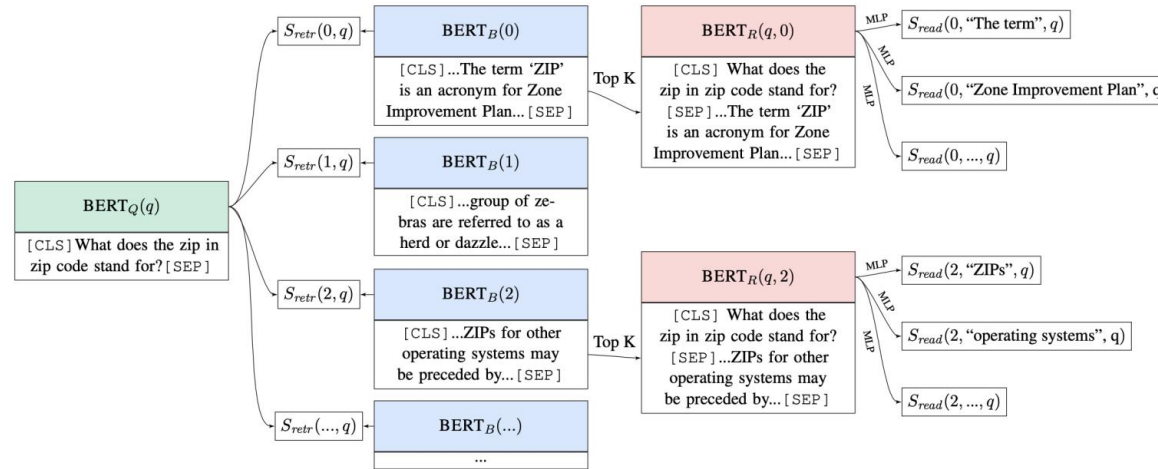


Image credit: (Lee et al., 2019)

# Question Answering

Beyond textual QA problems

오늘날에는 **unstructured text**에 기반한 질문에 답하고자 한다

## Knowledge based QA

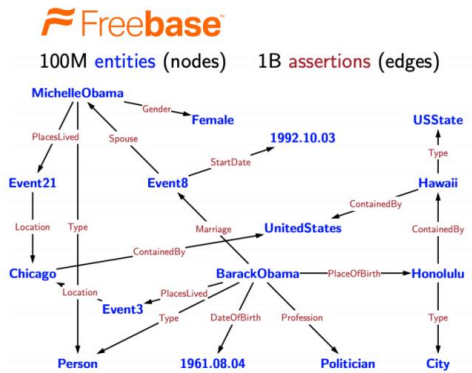


Image credit: Percy Liang

## Visual QA

Which states' capitals are also their largest cities by area?

semantic parsing

$\mu x. \text{Type.USState} \sqcap \text{Capital.argmax}(\text{Type.City} \sqcap \text{ContainedBy}.x, \text{Area})$

execute

Arizona, Hawaii, Idaho, Indiana, Iowa, Oklahoma, Utah



What color are her eyes?  
What is the mustache made of?



How many slices of pizza are there?  
Is this a vegetarian pizza?

(Antol et al., 2015): Visual Question Answering

# Reading Comprehension

What is reading comprehension?

Kannada language is the official language of Karnataka and spoken as a native language by about 66.54% of the people as of 2011. Other linguistic minorities in the state were Urdu (10.83%), Telugu language (5.84%), Tamil language (3.45%), Marathi language (3.38%), Hindi (3.3%), Tulu language (2.61%), Konkani language (1.29%), Malayalam (1.27%) and Kodava Takk (0.18%). In 2007 the state had a birth rate of 2.2%, a death rate of 0.7%, an infant mortality rate of 5.5% and a maternal mortality rate of 0.2%. The total fertility rate was 2.2.

Q: Which linguistic minority is larger, Hindi or Malayalam?

A: Hindi

## Information extraction

(Barack Obama, educated\_at, ?)

Question: Where did Barack Obama graduate from?

Passage: Obama was born in Honolulu, Hawaii. After graduating from Columbia University in 1983, he worked as a community organizer in Chicago.

(Levy et al., 2017)

## Semantic role labeling

UCD finished the 2006 championship as Dublin champions ,  
by beating St Vincents in the final .

finished

Who finished something? - UCD  
What did someone finish? - the 2006 championship  
What did someone finish something as? - Dublin champions  
How did someone finish something? - by beating St Vincents in the final

beating

Who beat someone? - UCD  
When did someone beat someone? - in the final  
Who did someone beat? - St Vincents

(He et al., 2015)

$$(P, Q) \rightarrow A$$

Text로 이루어진 문단을 이해하고  
해당 내용에 대한 질문에 답하자!

- ✓ 얼마나 컴퓨터가 사람의 언어를 잘 이해하는지 평가할 수 있는 testbed
- ✓ 많은 다른 NLP task도 reading comprehension 문제로 단순화 가능

# Reading Comprehension

Stanford question answering dataset (SQuAD)

(**passage**, **question**, **answer**) triple 100k개  
한계: 각 answer passage의 span

## Stanford Question Answering Dataset (SQuAD)

Private schools, also known as independent schools, non-governmental, or nonstate schools, are not administered by local, state or national governments; thus, they retain the right to select their students and are funded in whole or in part by charging their students tuition, rather than relying on mandatory taxation through public (government) funding; at some private schools students may be able to get a scholarship, which makes the cost cheaper, depending on a talent the student may have (e.g. sport scholarship, art scholarship, academic scholarship), financial need, or tax credit scholarships that might be available.

**Along with non-governmental and nonstate schools, what is another name for private schools?**

**Gold answers:** ① independent ② independent schools ③ independent schools

Estimated human performance  
EM = 82.3, F1 = 91.2

## Evaluation

### ✓ Exact Match (EM)

3개의 answer에 대해 span이 존재하는지에 따라 0/1 accuracy 값 획득

### ✓ F1

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

	Actual	
	True positive(TP)	False Positive(FP)
Predicted	False negative(FN)	True Negative(TN)



Q: What did Tesla do in December 1878?

A: {left Graz, left Graz, left Graz and severed all relations with his family}

Prediction: {left Graz and served}

Exact match:  $\max\{0, 0, 0\} = 0$

F1:  $\max\{0.67, 0.67, 0.61\} = 0.67$

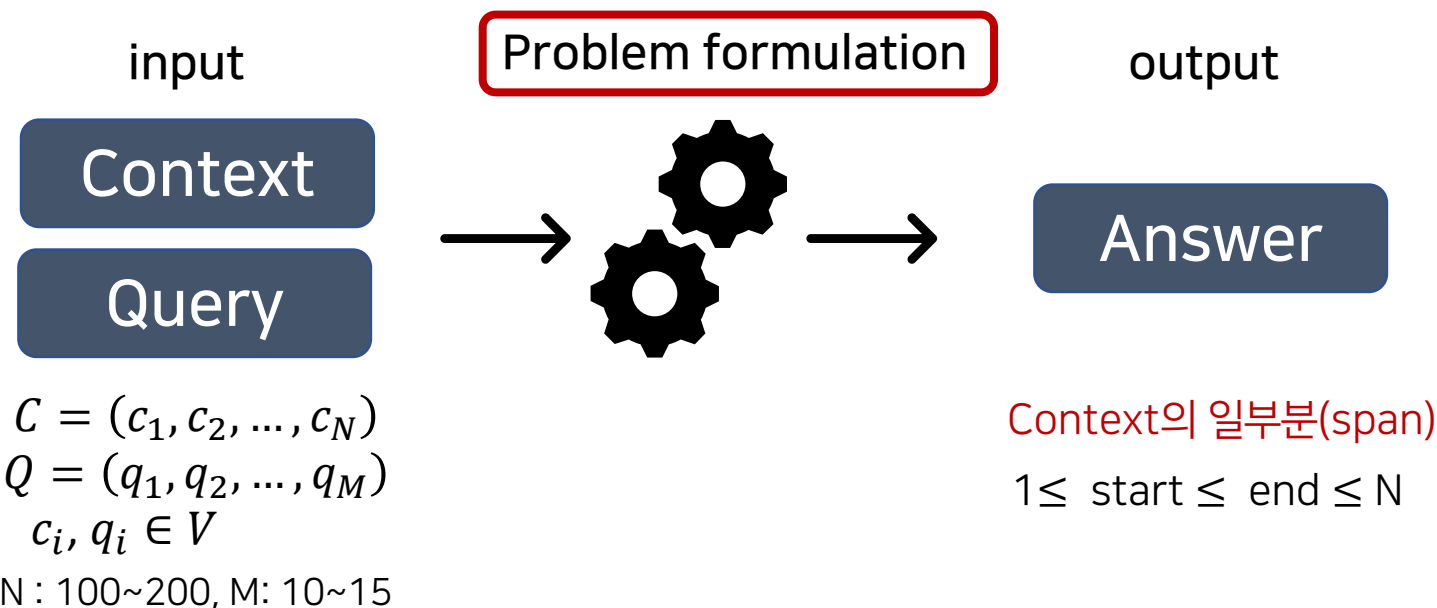
1. 예측된 answer을 각각의 gold answer와 비교  
(a, an, the, .(구두점) 제거)
2. Max score 계산
3. 모든 example에 대해 EM과 F1 평균

## 02

# Reading Comprehension

How can we build a model to solve SQuAD?

(passage, paragraph, context), (question query) 같은 의미로 사용

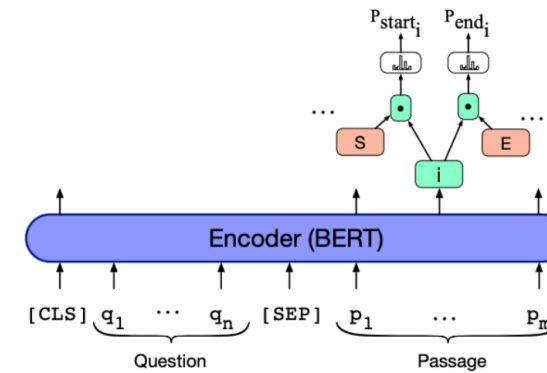
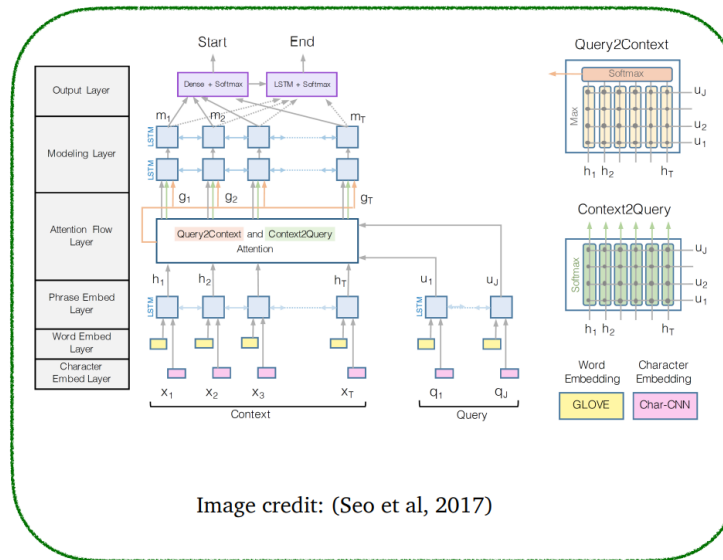


A family of LSTM-based models with attention (2016-2018)

Fine-tuning BERT-like models for reading comprehension(2019+)

# Reading Comprehension

LSTM-based vs BERT models



# Reading Comprehension

Recap : Seq2seq model with attention

## Machine Translation

- ✓ Source, target 문장
- ✓ Autoregressive decoder  
(word-by-word target 문장 생성)
- ✓ **Source** 문장의 어떤 단어가 현재의 **target** 단어와 가장 관련 있을까?

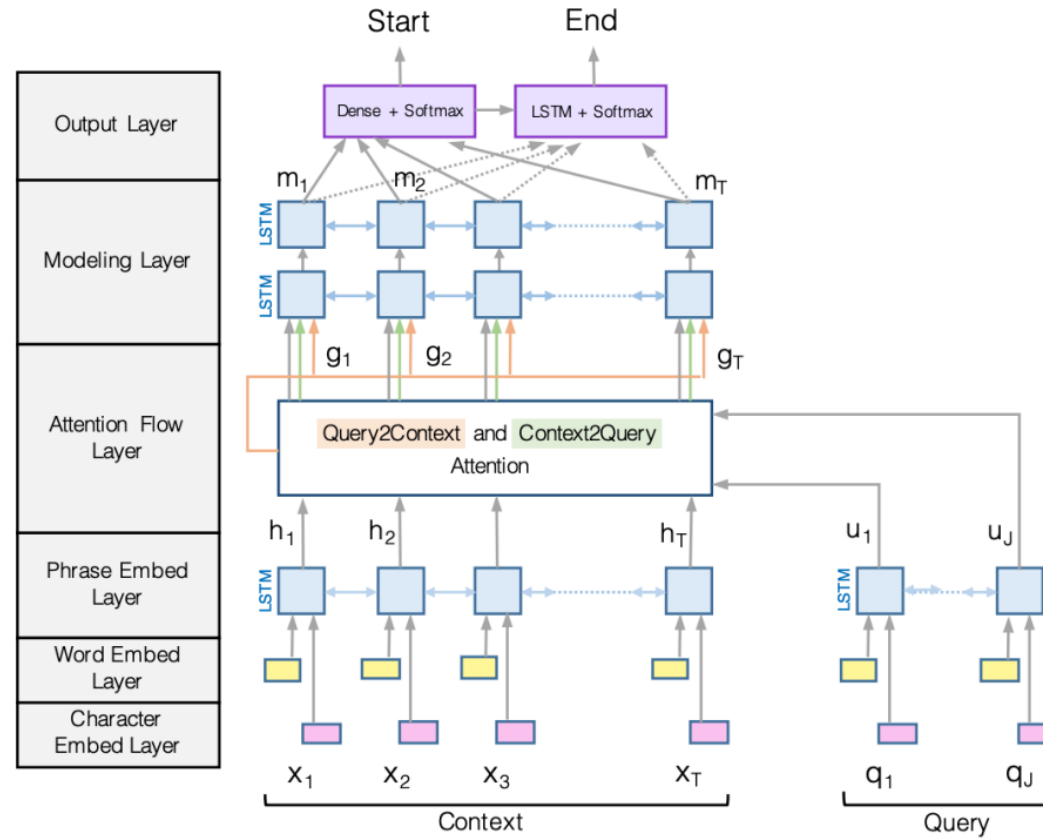
## Reading Comprehension

- ✓ Passage, question  
(길이는 다를 수 있다)
- ✓ Two classifier  
(정답의 start, end 위치만 예측)
- ✓ **Passage**의 어떤 단어들이 **question**(의 어떤 단어)와 가장 관련 있을까?

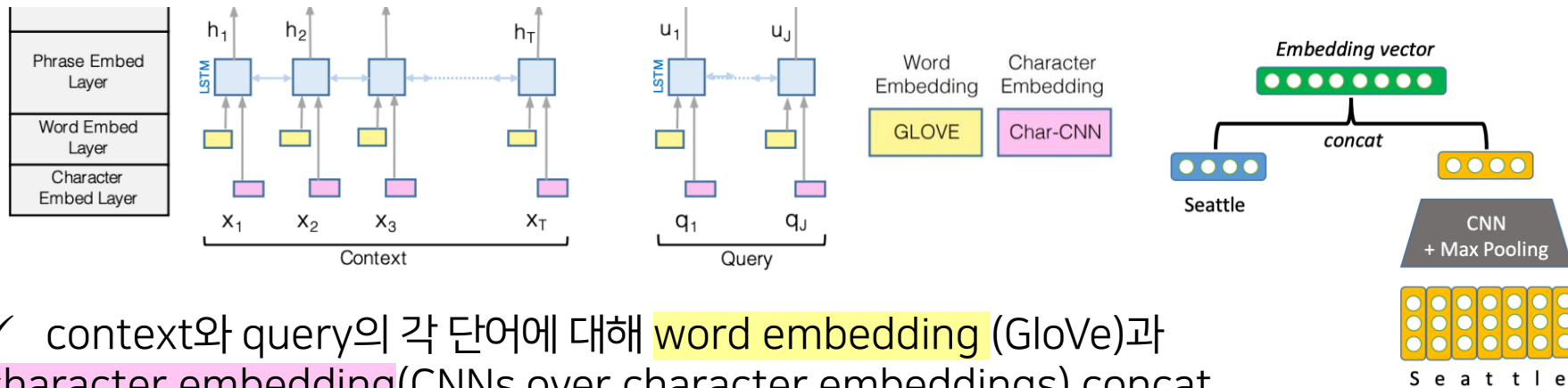
**Attention** is the key ingredient

# Reading Comprehension

The Bidirectional Attention Flow model(BiDAF)



(Seo et al., 2017): Bidirectional Attention Flow for Machine Comprehension



- ✓ context와 query의 각 단어에 대해 word embedding (GloVe)과 character embedding (CNNs over character embeddings) concat  

$$e(c_i) = f([GloVe(c_i); charEmb(c_i)]) \quad e(q_i) = f([GloVe(q_i); charEmb(q_i)])$$
 f: high-way networks omitted here

- ✓ contextual embedding 생성 위해 context와 query에 대해 각각 bidirectional LSTM 사용

$$\vec{c}_i = LSTM(\vec{c}_{i-1}, e(c_i)) \in \mathbb{R}^H$$

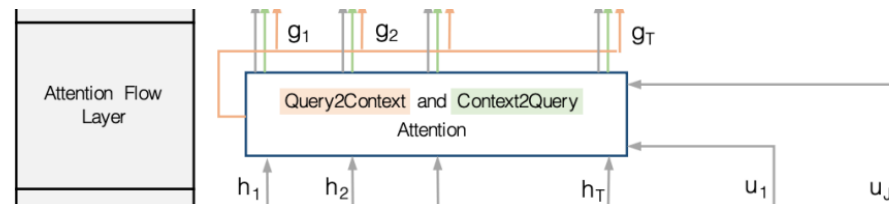
$$\tilde{c}_i = LSTM(\tilde{c}_{i+1}, e(c_i)) \in \mathbb{R}^H$$

$$c_i = [\vec{c}_i, \tilde{c}_i] \in \mathbb{R}^{2H}$$

$$\vec{q}_i = LSTM(\vec{q}_{i-1}, e(q_i)) \in \mathbb{R}^H$$

$$\tilde{q}_i = LSTM(\tilde{q}_{i+1}, e(q_i)) \in \mathbb{R}^H$$

$$q_i = [\vec{q}_i, \tilde{q}_i] \in \mathbb{R}^{2H}$$



## Query-to-context attention

query 단어 중 하나와 가장 관련  
있는 context 단어들 선택하기

While **Seattle**'s weather is very nice in summer, its weather is very rainy **in winter**, making it one of the most **gloomy cities** in the U.S. LA is ...

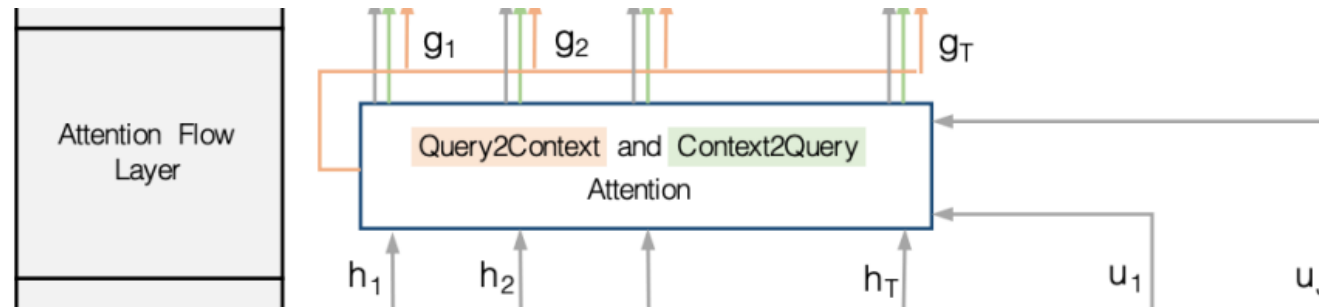
Q: Which city is gloomy in winter?

## Context-to-query attention

각각의 context 단어에 대해 가장  
유사한 query 단어 찾기

Q: Who leads the United States?

C: Barack Obama is the president of USA.



**Input** : query와 context의 contextual vector representation ( $c_i, q_j$ )

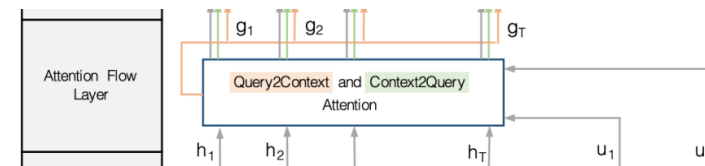
**Output** : context 단어들의 query-aware vector representation( $g_i$ ),  
이전 layer의 contextual embedding

1. 모든  $(c_i, q_j)$  쌍에 대해 유사도 (similarity score) 계산

$$S_{i,j} = w_{sim}^T [c_i; q_j; c_i \odot q_j] \in \mathbb{R} \quad w_{sim} \in \mathbb{R}^{6H}$$

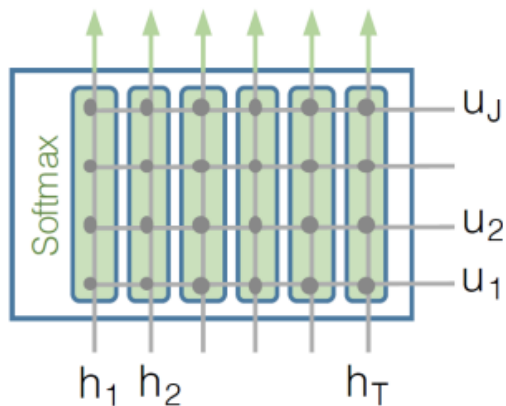
2. Context-to-query attention (질문의 어떤 단어들이  $c_i$ 와 관련있는지),  
Query-to-context attention (문단에서 어떤 단어들이 질문 단어들과  
관련있는지) 연산





## Context-to-query attention

각 context 단어에 어떤 query  
단어들이 더 관련있는지



$$\alpha_{i,j} = \text{softmax}_j(S_{i,j}) \in \mathbb{R}$$

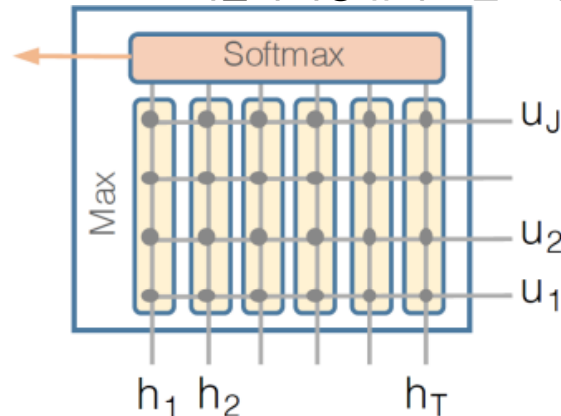
$$a_i = \sum_{j=1}^M \alpha_{i,j} q_j \in \mathbb{R}^{2H}$$

Final output

$$g_i = [c_i; a_i; c_i \odot a_i; c_i \odot b_i] \in \mathbb{R}^{8H}$$

## Query-to-context attention

하나의 query 단어에 대해 어떤  
context 단어들이 가장 유사도 높은지

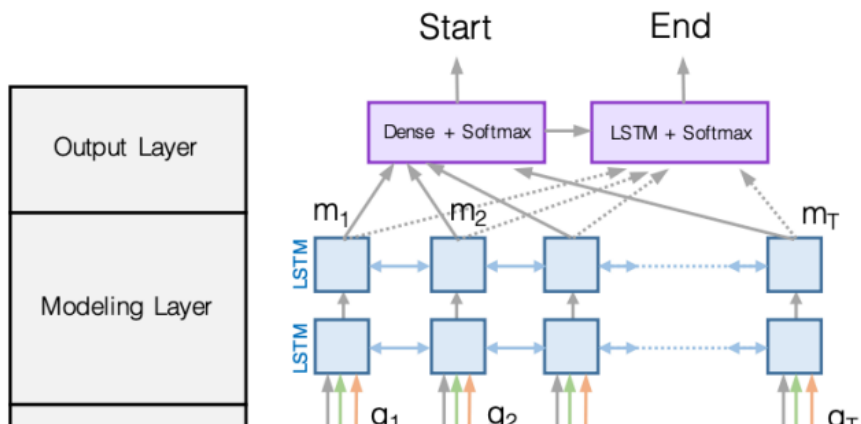


$$\beta_{i,j} = \text{softmax}_i(\max_{j=1}^M(S_{i,j})) \in \mathbb{R}^N$$

$$b_i = \sum_{i=1}^M \beta_i c_i \in \mathbb{R}^{2H}$$

# Reading Comprehension

BiDAF: Modeling and output layers



Final training loss

$$L = -\log p_{start}(s^*) - \log p_{end}(e^*)$$

✓ **Modeling layer** :  $g_i$ 를 또다른 bidirectional LSTM의 2개의 layer로 전달

-Attention layer : query와 context 사이의 interaction 모델링

-Modeling layer : context 단어들 사이의 interaction 모델링

$$m_i = BiLSTM(m_i) \in \mathbb{R}^{2H}$$

✓ **Output layer** : start, end 위치를 예측하는 classifier

$$p_{start} = \text{softmax}(w_{start}^T [g_i; m_i]) \quad p_{end} = \text{softmax}(w_{end}^T [g_i; m'_i])$$

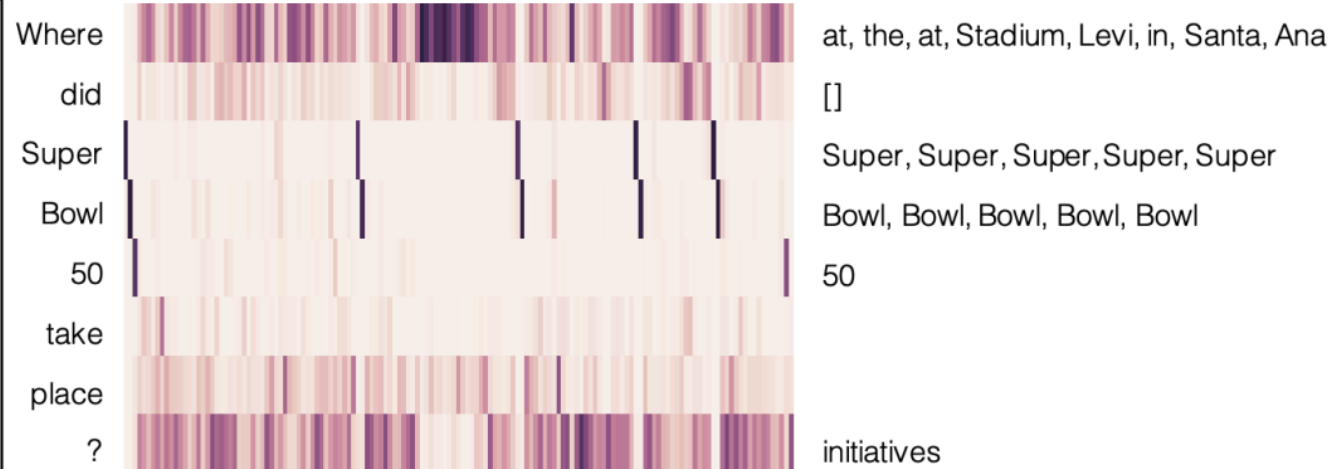
$$m'_i = BiLSTM(m_i) \in \mathbb{R}^{2H} \quad w_{start}, w_{end} \in \mathbb{R}^{10H}$$

	EM	F1
No char embedding	65.0	75.4
No word embedding	55.5	66.8
No C2Q attention	57.2	67.7
No Q2C attention	63.6	73.7
Dynamic attention	63.5	73.6
BiDAF (single)	67.7	77.3
BiDAF (ensemble)	72.6	80.7

	Published <sup>12</sup>	LeaderBoard <sup>13</sup>
Single Model	EM / F1	EM / F1
LR Baseline (Rajpurkar et al., 2016)	40.4 / 51.0	40.4 / 51.0
Dynamic Chunk Reader (Yu et al., 2016)	62.5 / 71.0	62.5 / 71.0
Match-LSTM with Ans-Ptr (Wang & Jiang, 2016)	64.7 / 73.7	64.7 / 73.7
Multi-Perspective Matching (Wang et al., 2016)	65.5 / 75.1	70.4 / 78.8
Dynamic Coattention Networks (Xiong et al., 2016)	66.2 / 75.9	66.2 / 75.9
FastQA (Weissenborn et al., 2017)	68.4 / 77.1	68.4 / 77.1
BiDAF (Seo et al., 2016)	68.0 / 77.3	68.0 / 77.3
SED (Liu et al., 2017a)	68.1 / 77.5	68.5 / 78.0
RaSoR (Lee et al., 2016)	70.8 / 78.7	69.6 / 77.7
FastQAExt (Weissenborn et al., 2017)	70.8 / 78.9	70.8 / 78.9
ReasonNet (Shen et al., 2017b)	69.1 / 78.9	70.6 / 79.4
Document Reader (Chen et al., 2017)	70.0 / 79.0	70.7 / 79.4
Ruminating Reader (Gong & Bowman, 2017)	70.6 / 79.5	70.6 / 79.5
jNet (Zhang et al., 2017)	70.6 / 79.8	70.6 / 79.8
Conductor-net	N/A	72.6 / 81.4
Interactive AoA Reader (Cui et al., 2017)	N/A	73.6 / 81.9
Reg-RaSoR	N/A	75.8 / 83.3
DCN+	N/A	74.9 / 82.8
AIR-FusionNet	N/A	76.0 / 83.9
R-Net (Wang et al., 2017)	72.3 / 80.7	76.5 / 84.3
BiDAF + Self Attention + ELMo	N/A	<b>77.9 / 85.3</b>
Reinforced Mnemonic Reader (Hu et al., 2017)	73.2 / 81.8	73.2 / 81.8

## Attention score

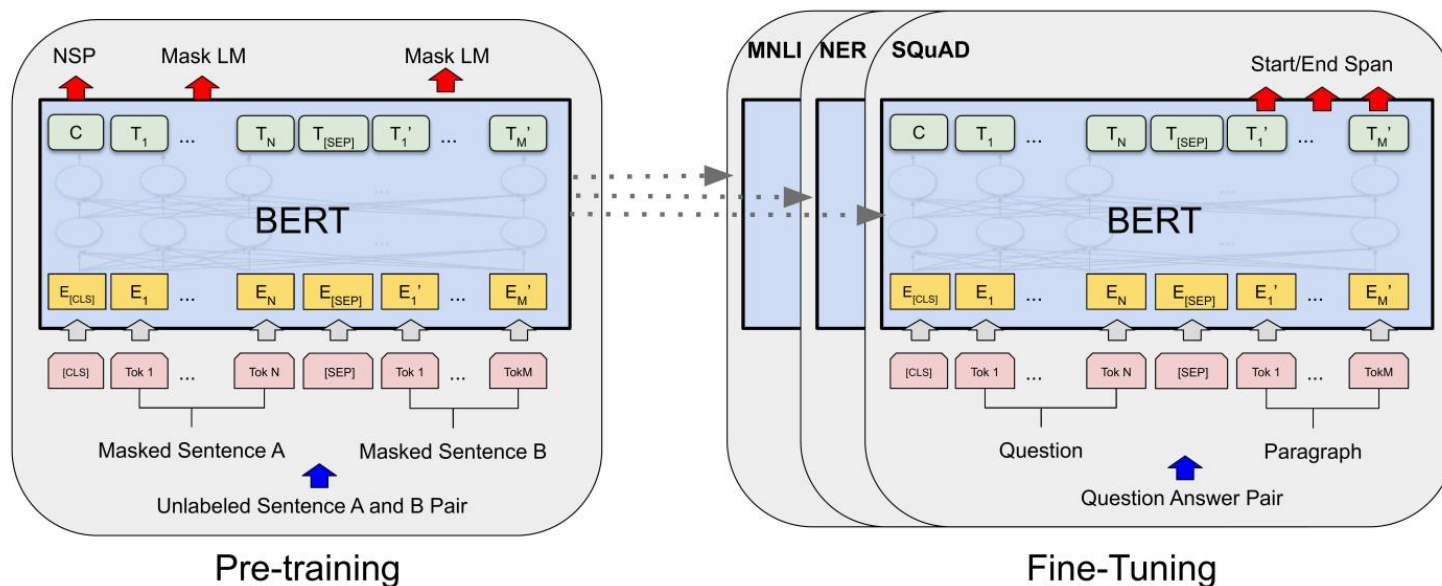
Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season . The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title . The game was played on February 7 , 2016 , **at Levi 's Stadium in the San Francisco Bay Area at Santa Clara , California** . As this was the 50th Super Bowl , the league emphasized the " golden anniversary " with various gold-themed initiatives , as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals ( under which the game would have been known as " Super Bowl L " ) , so that the logo could prominently feature the Arabic numerals 50 .



# Reading Comprehension

BERT for reading comprehension

**BERT란?** 대량의 text(Wikipedia + BooksCorpus) 에 pre-train된 deep bidirectional Transformer encoder  
 pre-train 1. Masked language model (MLM) 2. Next sentence prediction (NSP)

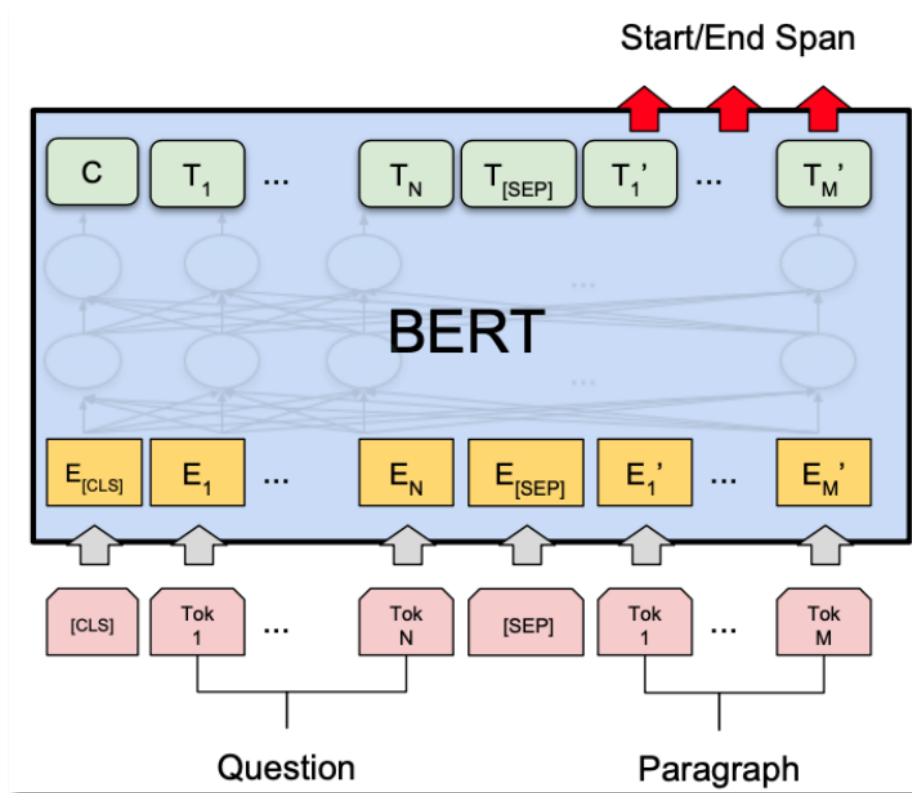


$BERT_{base}$  : 12개의 layer, 110M parameters

$BERT_{large}$  : 24개의 layer, 330M parameters

# Reading Comprehension

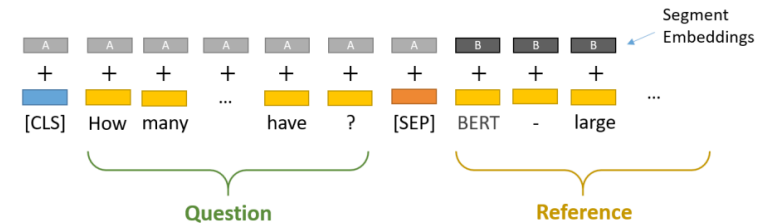
BERT for reading comprehension



Question = Segment A

Passage = Segment B

Answer = predicting two endpoints in segment B



**Question:** How many parameters does BERT-large have?

**Reference Text:** BERT-large is really big... it has 24 layers and an embedding size of 1,024, for a total of 340M parameters! Altogether it is 1.34GB, so expect it to take a couple minutes to download to your Colab instance.

Final training loss

$$L = -\log p_{start}(s^*) - \log p_{end}(e^*)$$

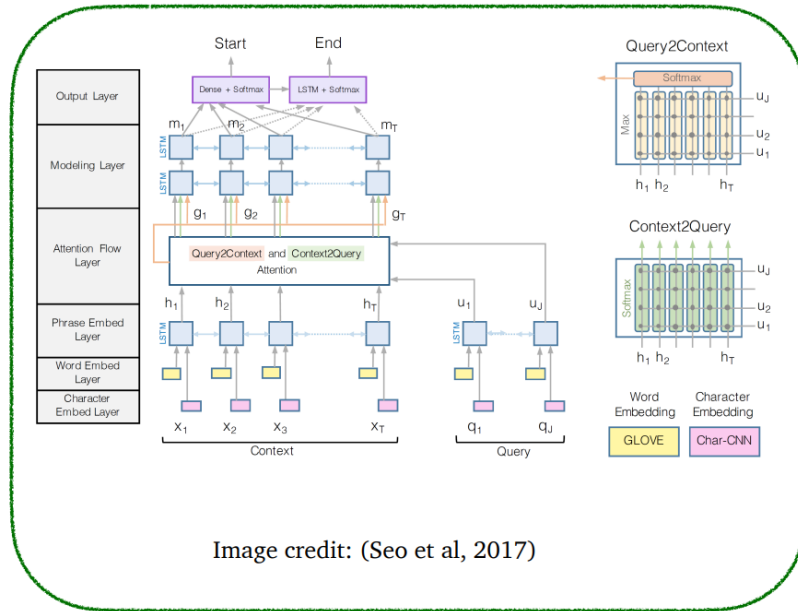
$$p_{start}(i) = \text{softmax}_i(w_{start}^T H)$$

$$p_{end}(i) = \text{softmax}_i(w_{end}^T H)$$

Where  $H = [h_1, h_2, \dots, h_N]$  는 BERT에 의해 반환되는 paragraph의 hidden vector

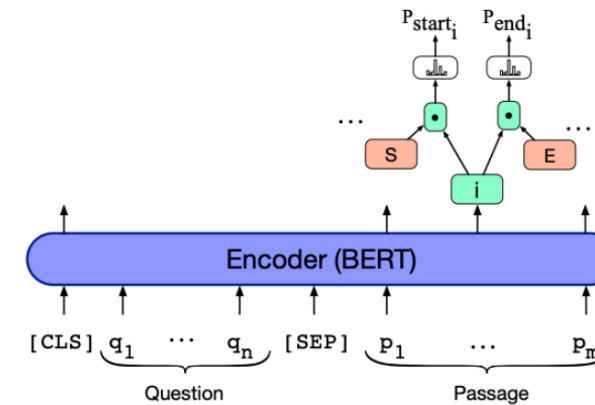
# Reading Comprehension

Comparisons between BiDAF and BERT models



## BiDAF

- ✓ ~2.5M parameters
- ✓ 다수의 bidirectional LSTM 위에 build
- ✓ Only built on top of GloVe



## BERT models

- ✓ 110M or 330M parameters
- ✓ Transformer 위에 build
- ✓ Pre-trained



- ✓ Question과 passage 사이의 interaction model
- ✓ BERT는 question과 passage의 **concatenation** 사이에서 self-attention 사용 =  $\text{attention}(P, P) + \text{attention}(P, Q) + \text{attention}(Q, P) + \text{attention}(Q, Q)$
- ✓ BiDAF에 passage에 대한 self-attention layer ( $\text{attention}(P, P)$ ) 추가하면 성능이 좋아진다

✓ ~2.5M parameters

✓ 다수의 bidirectional LSTM 위에 build

✓ Only built on top of GloVe

✓ 110M or 330M parameters

✓ Built on top of Transformers

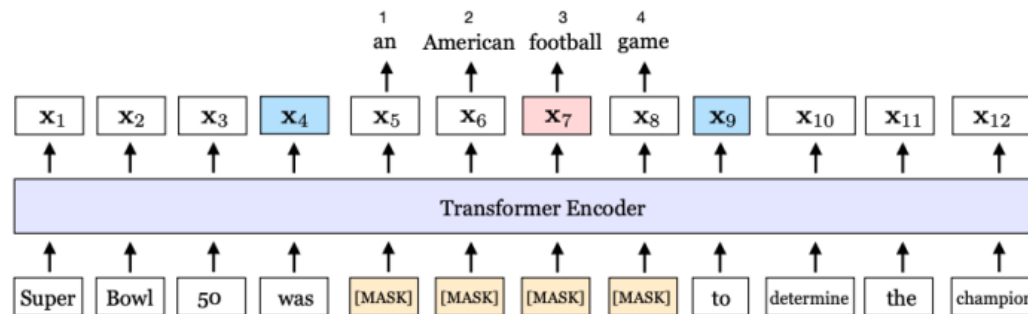
✓ Pre-trained



# Reading Comprehension

Can we design better pre-training objectives?

$$\begin{aligned}\mathcal{L}(\text{football}) &= \mathcal{L}_{\text{MLM}}(\text{football}) + \mathcal{L}_{\text{SBO}}(\text{football}) \\ &= -\log P(\text{football} \mid \mathbf{x}_7) - \log P(\text{football} \mid \mathbf{x}_4, \mathbf{x}_9, \mathbf{p}_3)\end{aligned}$$



- ✓ 15%의 랜덤 단어 대신 **인접한 단어의 span**을 마스킹하자
- ✓ endpoint 사이에 있는 마스크된 **단어들을 예측**하기 위해 span의 2개의 **endpoint** 사용 하자
- = 2개의 endpoint에 span의 정보를 압축하여 넣자

(Joshi & Chen et al., 2020): SpanBERT: Improving Pre-training by Representing and Predicting Spans

# Reading Comprehension

Is reading comprehension solved?

**Article:** Super Bowl 50

**Paragraph:** “Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.”

**Question:** “What is the name of the quarterback who was 38 in Super Bowl XXXIII?”

**Original Prediction:** John Elway

**Prediction under adversary:** Jeff Dean

Adversarial distracting sentence

SQuAD에 대해서는 이미 인간보다 뛰어남!  
그렇다면 **reading comprehension**은 다  
해결된 문제라 볼 수 있을까? 아니다!

## 문제점

- ✓ Adversarial example에 대해 낮은 성능을 보임
- ✓ Out-of-domain distribution의 example에 대해 낮은 성능

	Match Single	Match Ens.	BiDAF Single	BiDAF Ens.
Original	71.4	75.4	75.5	80.0
ADDSSENT	27.3	29.4	34.3	34.2
ADDONESSENT	39.0	41.8	45.7	46.9
ADDANY	7.6	11.7	4.8	2.7
ADDCOMMON	38.9	51.0	41.7	52.6

(Jia and Liang, 2017): Adversarial Examples for Evaluating Reading Comprehension Systems

한 데이터셋에 대해 train된 시스템들은 다른 데이터셋으로 generalize 못함

		Evaluated on				
		SQuAD	TriviaQA	NQ	QuAC	NewsQA
Fine-tuned on	SQuAD	<b>75.6</b>	46.7	48.7	20.2	41.1
	TriviaQA	49.8	<b>58.7</b>	42.1	20.4	10.5
	NQ	53.5	46.3	<b>73.5</b>	21.6	24.7
	QuAC	39.4	33.1	33.8	<b>33.3</b>	13.8
	NewsQA	52.1	38.4	41.7	20.4	<b>60.1</b>

(Sen and Saffari, 2020): What do Models Learn from Question Answering Datasets?

# Reading Comprehension

Is reading comprehension solved?

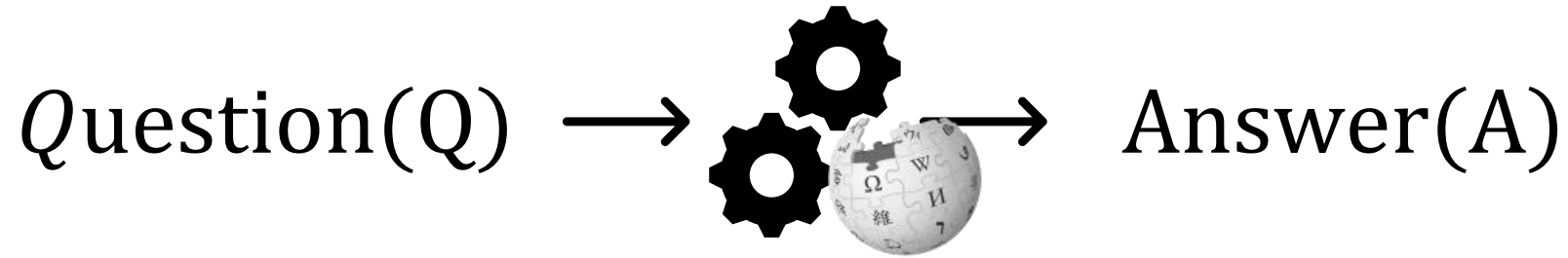
## BERT-large model trained on SQuAD

Temporal	<b>MFT:</b> change in one person only	41.5	C: Both Luke and Abigail were writers, but there was a change in Abigail, who is now a model. Q: Who is a model? A: Abigail : Abigail were writers, but there was a change in Abigail
	<b>MFT:</b> Understanding before/after, last/first	82.9	C: Logan became a farmer before Danielle did. Q: Who became a farmer last? A: Danielle : Logan
Neg.	<b>MFT:</b> Context has negation	67.5	C: Aaron is not a writer. Rebecca is. Q: Who is a writer? A: Rebecca : Aaron
	<b>MFT:</b> Q has negation, C does not	100.0	C: Aaron is an editor. Mark is an actor. Q: Who is not an actor? A: Aaron : Mark
Coref.	<b>MFT:</b> Simple coreference, he/she.	100.0	C: Melissa and Antonio are friends. He is a journalist, and she is an adviser. Q: Who is a journalist? A: Antonio : Melissa
	<b>MFT:</b> Simple coreference, his/her.	100.0	C: Victoria and Alex are friends. Her mom is an agent Q: Whose mom is an agent? A: Victoria : Alex
	<b>MFT:</b> former/latter	100.0	C: Kimberly and Jennifer are friends. The former is a teacher Q: Who is a teacher? A: Kimberly : Jennifer
SRL	<b>MFT:</b> subject/object distinction	60.8	C: Richard bothers Elizabeth. Q: Who is bothered? A: Elizabeth : Richard
	<b>MFT:</b> subj/obj distinction with 3 agents	95.7	C: Jose hates Lisa. Kevin is hated by Lisa. Q: Who hates Kevin? A: Lisa : Jose

(Ribeiro et al., 2020): Beyond Accuracy: Behavioral Testing of NLP Models with Checklist

## Open-domain question answering

Open-domain question answering



Open- domain vs closed-domain?

- ✓ passage가 주어져 있다고 가정하지 않음
- ✓ 대신 다량의 document에 접근 가능 (ex: Wikipedia)
- ✓ 정답이 어디에 위치해 있는지 모름
- ✓ Challenging but practical

어떠한 **open-domain** 질문에도 답하자!

# Open-domain question answering

Retriever-reader framework

## Open-domain QA

SQuAD, TREC, WebQuestions, WikiMovies

Q. How many Warsaw's inhabitants spoke Polish in 1933?



WIKIPEDIA  
The Free Encyclopedia

질문 관련 문서 검색

Document  
Retriever

Warsaw

From Wikipedia, the free encyclopedia

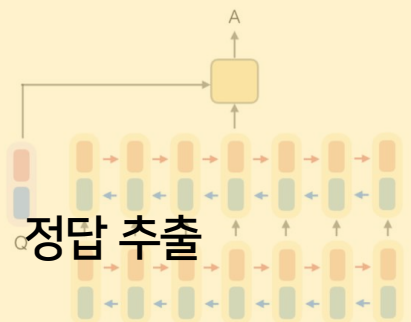
This article is about the Polish capital. For other uses, see Warsaw (disambiguation).  
 "Warsaw" redirects here. For other uses, see Warsaw (disambiguation).  
 "City of Warsaw" redirects here. For the Second World War fighter squadron, see No. 21st Polish Fighter Squadron, 1939. See Adamowski Brothers.

**Warsaw** (Polish: *Warszawa* [varˈʂava]; see also other names) is the capital and largest city of Poland. It stands on the Vistula River in east-central Poland, roughly 260 kilometres (160 mi) from the Baltic Sea and 300 kilometres (190 mi) from the Carpathian Mountains. Its population is estimated at 1,790 million residents within a greater metropolitan area of 2,105 million residents, which makes Warsaw the 9th-most populous capital city in the European Union.<sup>[2019]</sup> The city limits cover 516.9 square kilometres (199.6 sq mi), while the metropolitan area covers 6,100.43 square kilometres (2,357.07 sq mi).<sup>[2]</sup>

In 2012 the Economist Intelligence Unit ranked Warsaw as the 32nd most livable city in the world.<sup>[6]</sup> It was also ranked as one of the most livable cities in Central Europe. Today Warsaw is considered an "Alpha+" global city, a major international tourist destination and a significant cultural, political and economic hub.<sup>[7][8]</sup> Warsaw's economy, by a wide variety of industries, is characterised by FMCG manufacturing, metal processing, steel and electronic manufacturing and food processing. The city is a significant centre of research and development, R&D, IT, as well as of the Polish media industry. The Warsaw Stock Exchange is one of the largest and most important in Central and Eastern Europe.<sup>[19]</sup> France, the European Union agency for external border security, has its headquarters in Warsaw. It has been said that Warsaw, together with Frankfurt, London, Paris and Barcelona is one of the cities with the highest number of skyscrapers in the European Union.<sup>[19]</sup> Warsaw has also been called "Eastern Europe's chic cultural capital with thriving art and club scenes and serious restaurants".<sup>[19]</sup>

Document  
Reader

833,500



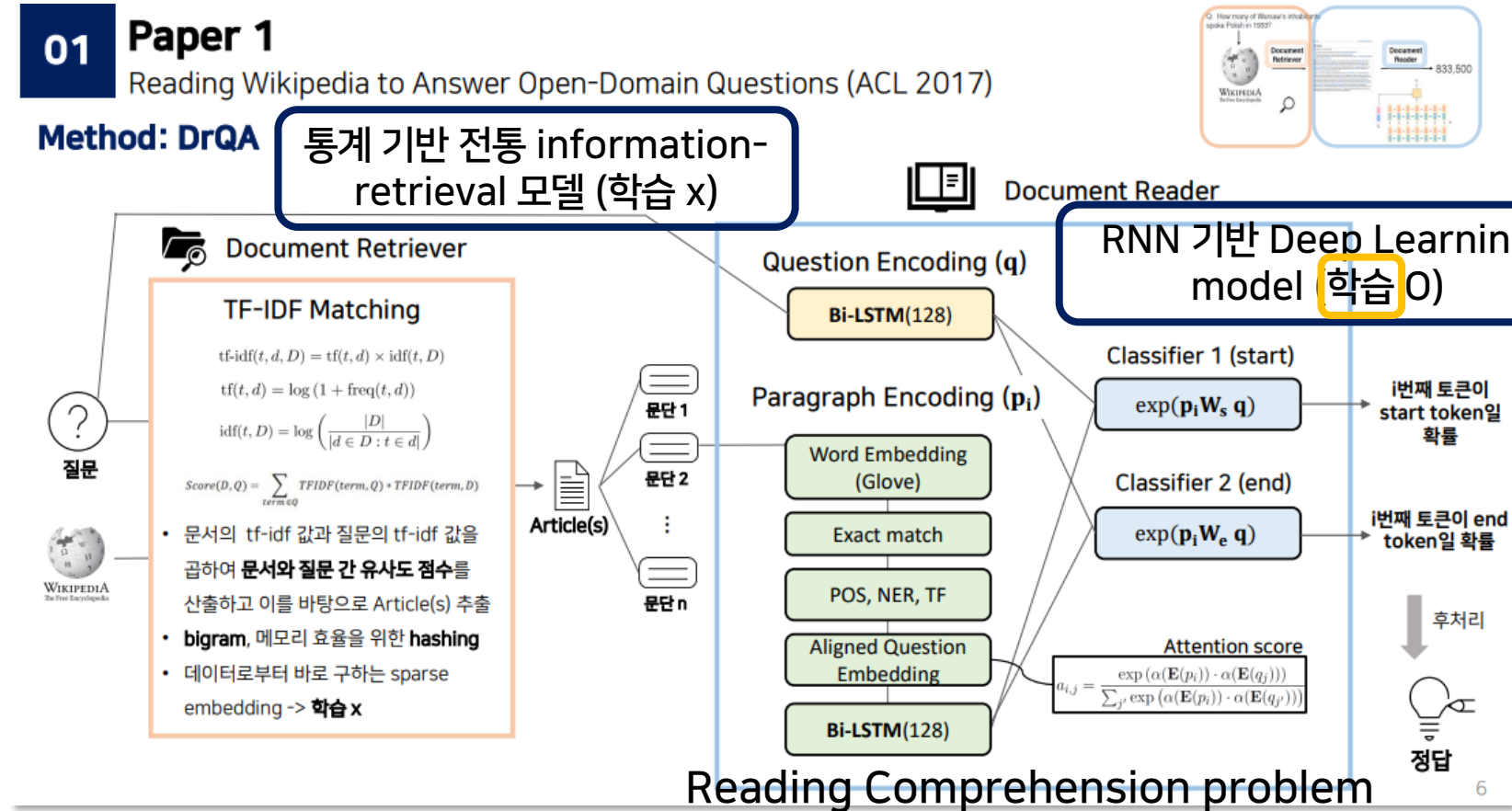
Chen et al., 2017. Reading Wikipedia to Answer Open-domain Questions

## 01 Paper 1

Reading Wikipedia to Answer Open-Domain Questions (ACL 2017)

Method: DrQA

통계 기반 전통 information-retrieval 모델 (학습 x)

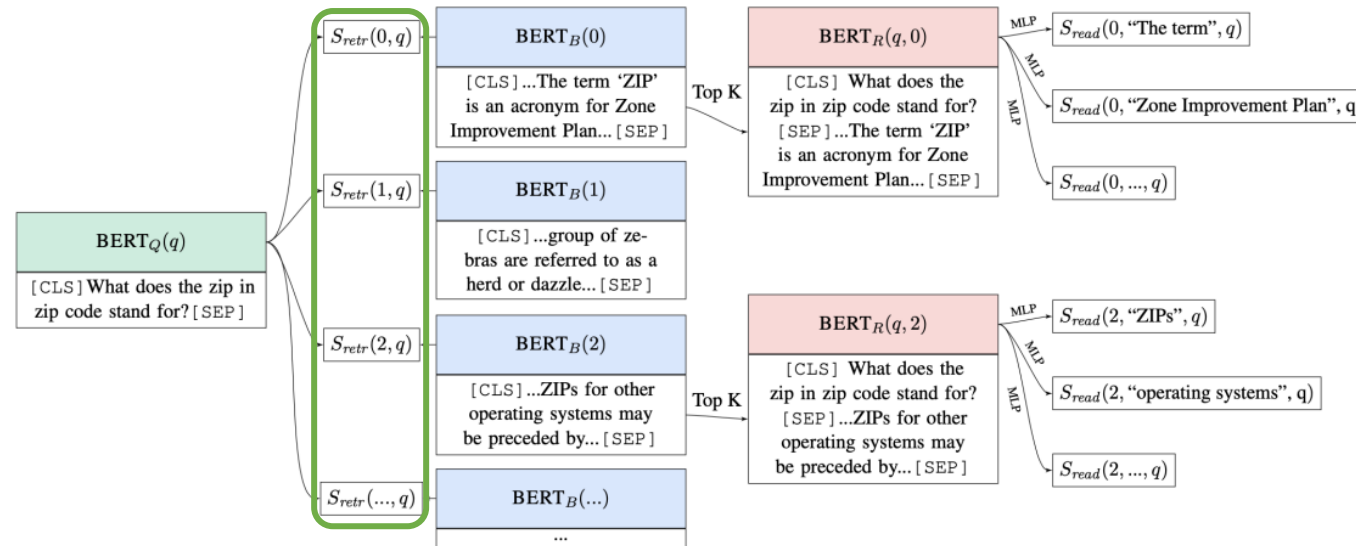


고유경 석사과정 Open-Domain Question Answering Paper Review #1

# Open-domain question answering

We can train the retriever too

Joint training of retriever and reader



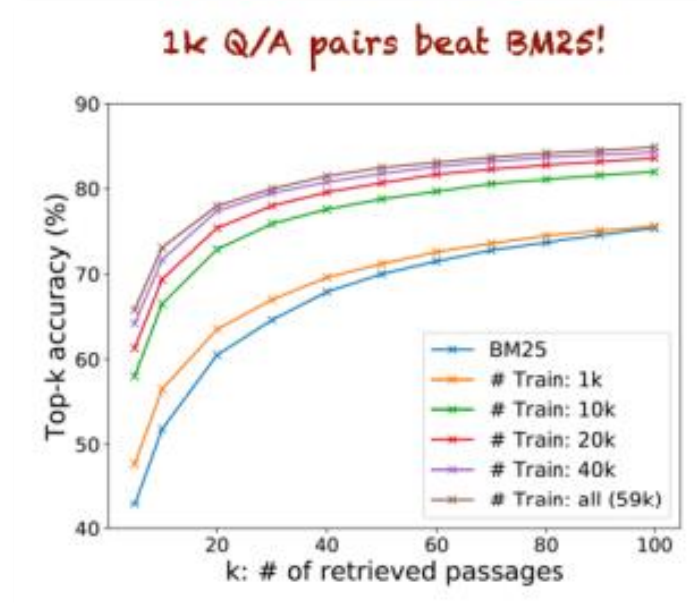
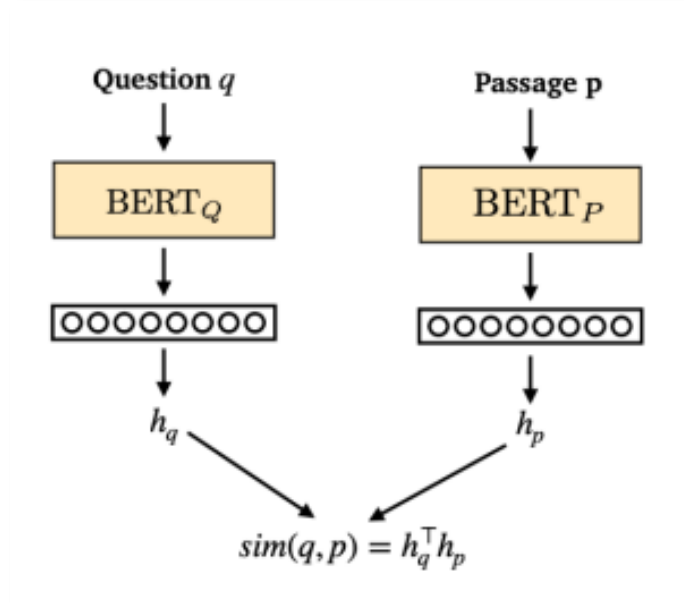
- ✓ question과 passage는 BERT를 이용해 encode될 수 있음
- ✓ Retrieval score = question representation과 passage representation 사이의 dot product
- ✓ BUT passage 수가 많을 때는 모델링이 쉽지 않음

Lee et al., 2019. Latent Retrieval for Weakly Supervised Open Domain Question Answering



## Open-domain question answering

We can train the retriever too



Dense passage retrieval (DPR)

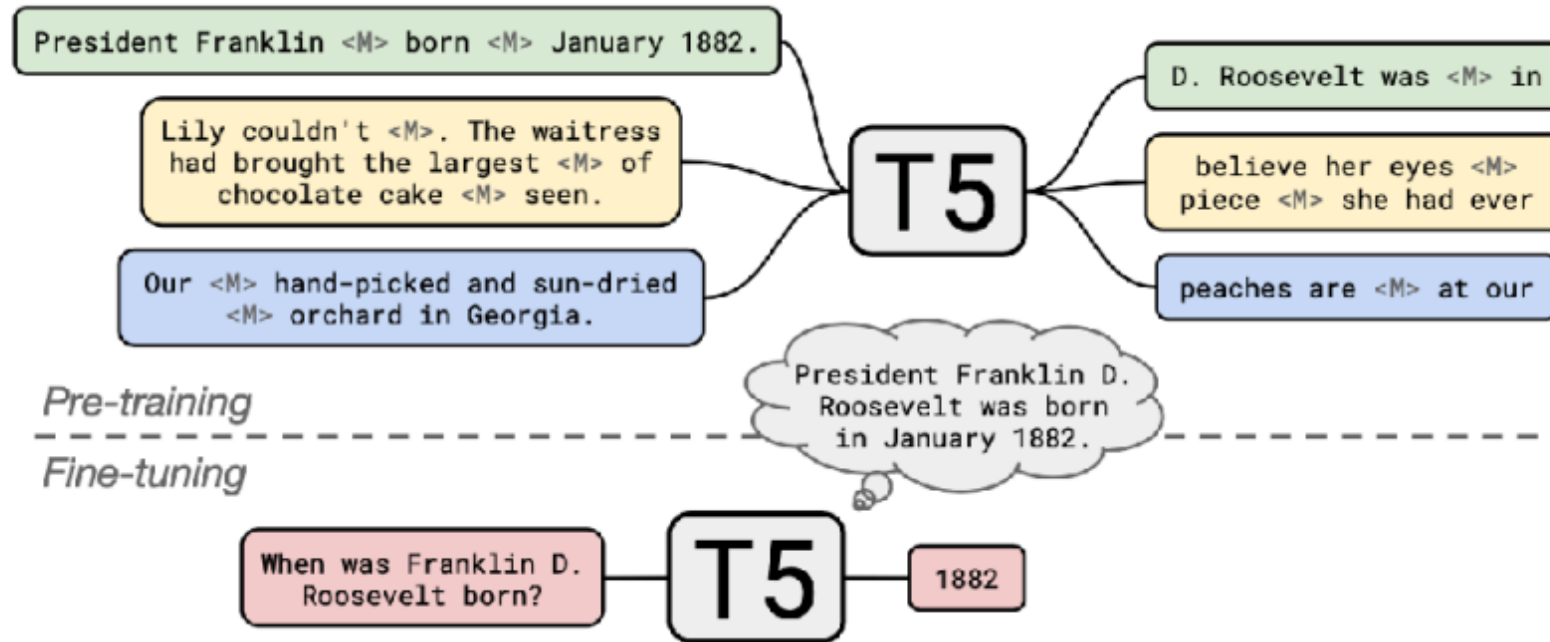
(Q,A) 쌍 이용해 retriever 학습 가능

Karpukhin et al., 2020. Dense Passage Retrieval for Open-Domain Question Answering

## 03

# Open-domain question answering

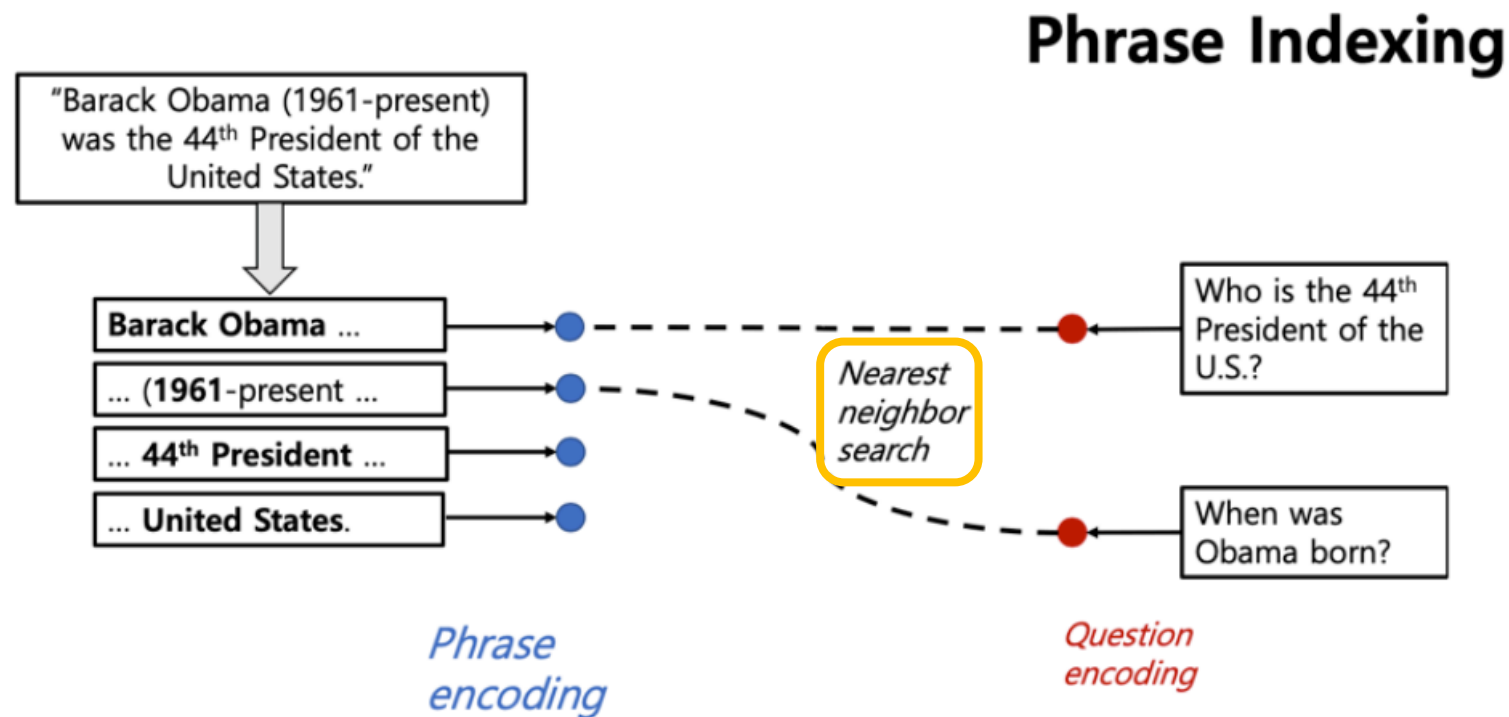
Without an explicit retrieval stage



Roberts et al., 2020. How Much Knowledge Can You Pack Into the Parameters of a Language Model?

## Open-domain (textual) question answering

Maybe the reader model is not necessary too!



Seo et al., 2019. Real-Time Open-Domain Question Answering with Dense-Sparse Phrase Index

Lee et al., 2020. Learning Dense Representations of Phrases at Scale 49 DensePhrases: Dem

**감사합니다**

00

# Appendix

000