JARDIS
Question Answering System:
Just A Rather Deep Intelligent System
TA: Zakaria Khiati
Team members: 유재준, 손규빈, 정유진, 류영표, 이현호

Our project scope

	Training				Implementation			
	week1	week2	week3	week4	week5	week6	week7	week8
Word embedding	■	■						
Sequence embedding		■	■					
Question Answering			■	■				
Memory Network				■	■			
Goal setting	■	■	■	■	■			
Dataset			■	■	■	■		
Implementation		■	■	■	■	■	■	■

While 공부 do 코딩

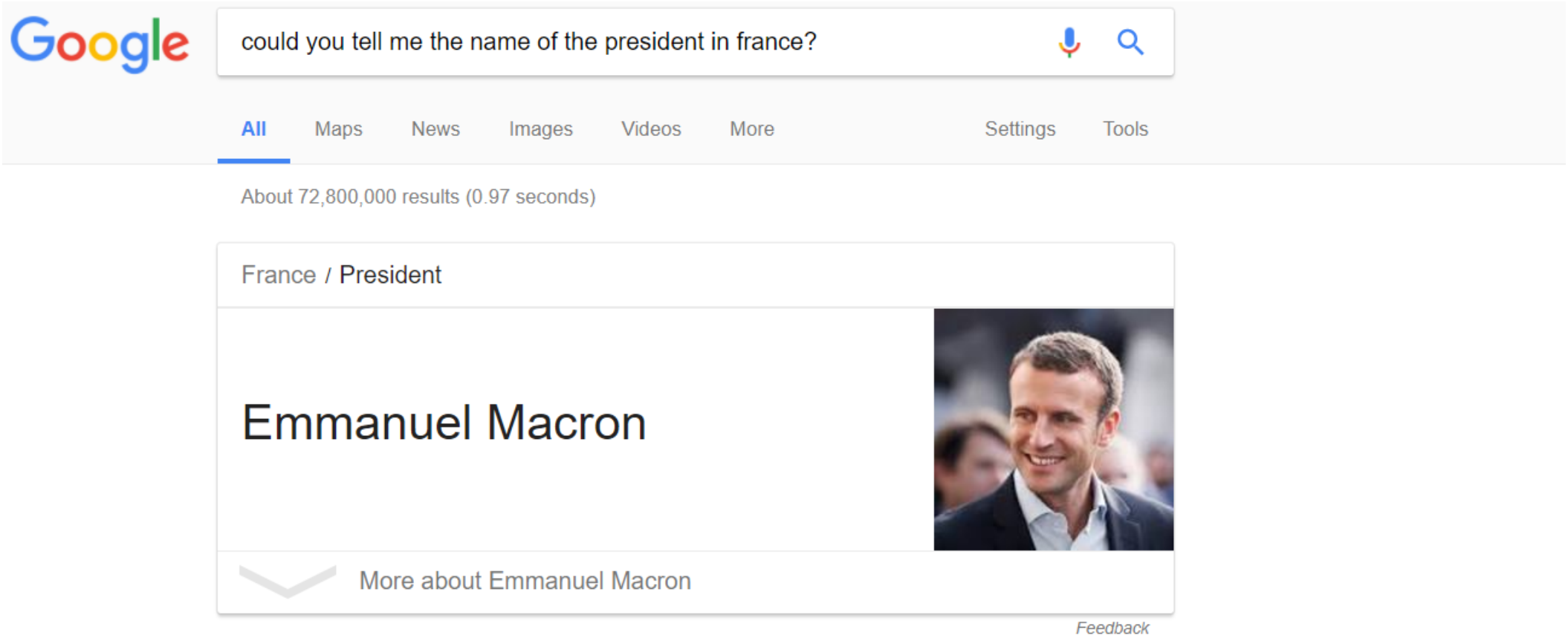
Our project scope

1. **Build a working QA system from a scratch**
   - Problem setup
   - Understanding the whole system from data to algorithm
   - Knowing the current state of the art
2. **Attack more advanced application or problem set**
   - SQuAD
   - Knowledge based + DL QA system

Question Answering (QA) System

Could you tell me the name of the president in France?

Question Answering (QA) System

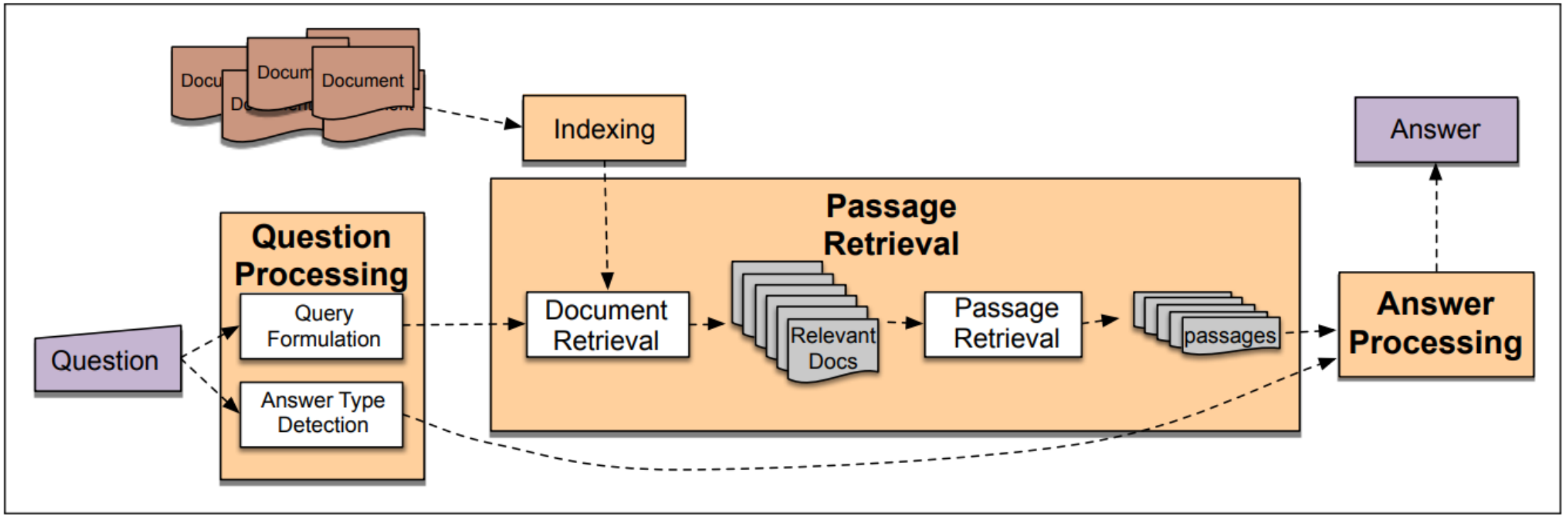


Question Answering (QA) System

Question answering (QA) is a computer science discipline within the fields of **information retrieva**l and **natural language processing (NLP)**, which is concerned with building systems that automatically answer questions posed by humans in a **natural language**.

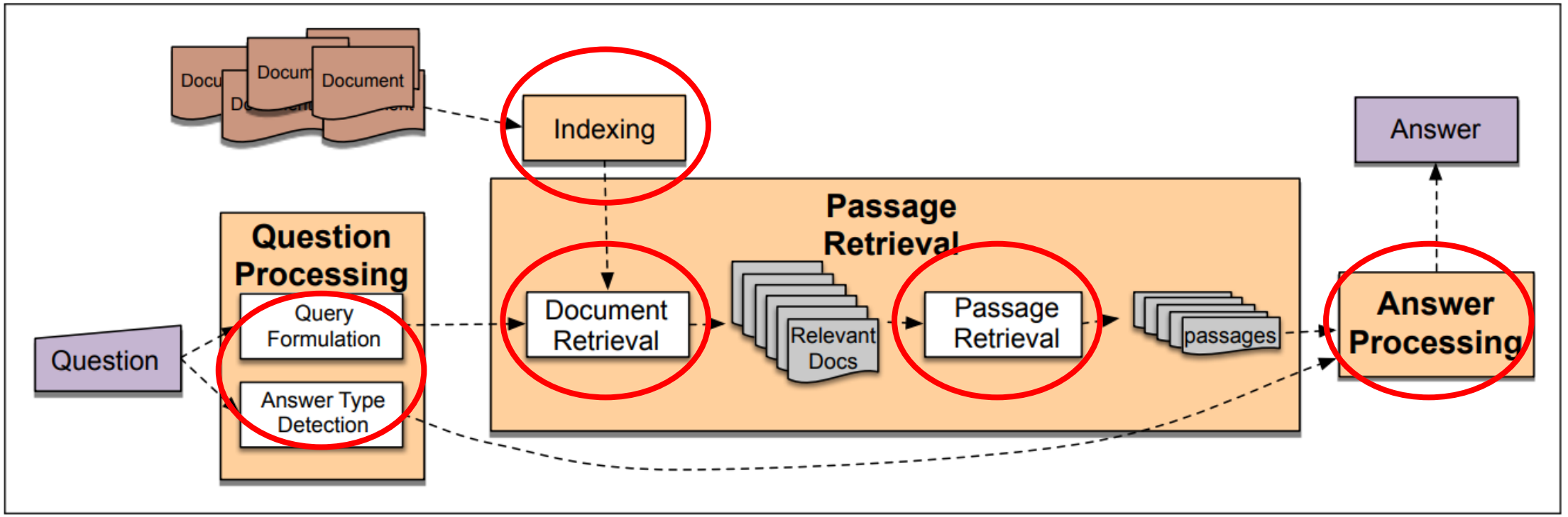
- Closed-domain
- Open-domain

Question Answering (QA) System



Can tackle open domain QA problem.
: very powerful and general

Question Answering (QA) System



Every module needs a separate treatment of its own
: Complex, needs for heavy domain expertise

Question Answering (QA) System

Deep learning !

- End-to-end
- Fast and efficient (if it can be trained)
- Easy to start for newbies (well…at least easy to try)
- Better performance? (maybe… if we have enough data and good structure based on strong domain knowledge)

Question Answering (QA) System

Deep learning !

- End-to-end
- Fast and efficient (if it can be trained)
- Easy to start for newbies (well…at least easy to try)
- Better performance? (maybe… if we have enough data and good structure based on strong domain knowledge)

Question Answering (QA) System

Deep learning !

- End-to-end
- Fast and efficient (if it can be trained)
- Easy to start for newbies (well…at least easy to try)
- Better performance? (maybe… if we have enough data and good structure based on strong domain knowledge)

Question Answering (QA) System

Deep learning !

- End-to-end
- Fast and efficient (if it can be trained)
- Easy to start for newbies (well…at least easy to try)
- Better performance? (maybe… providing another venue for improvements)

How to QA? – Problem Setup

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *arXiv preprint arXiv:1409.0575*. [web]

How to QA? – Problem Setup

Facebook: bAbI

Two supporting facts **story**

1 Mary got the milk there.
2 John moved to the bedroom.
3 Sandra went back to the kitchen.
4 Mary travelled to the hallway.

5 Where is the milk? **question**

answer

hallway | 1 4 **supporting line**

SQuAD

The Stanford Question Answering Dataset

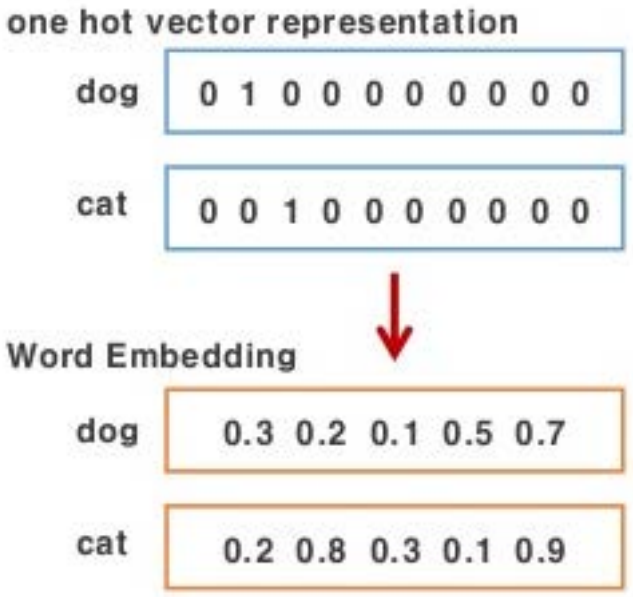
The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

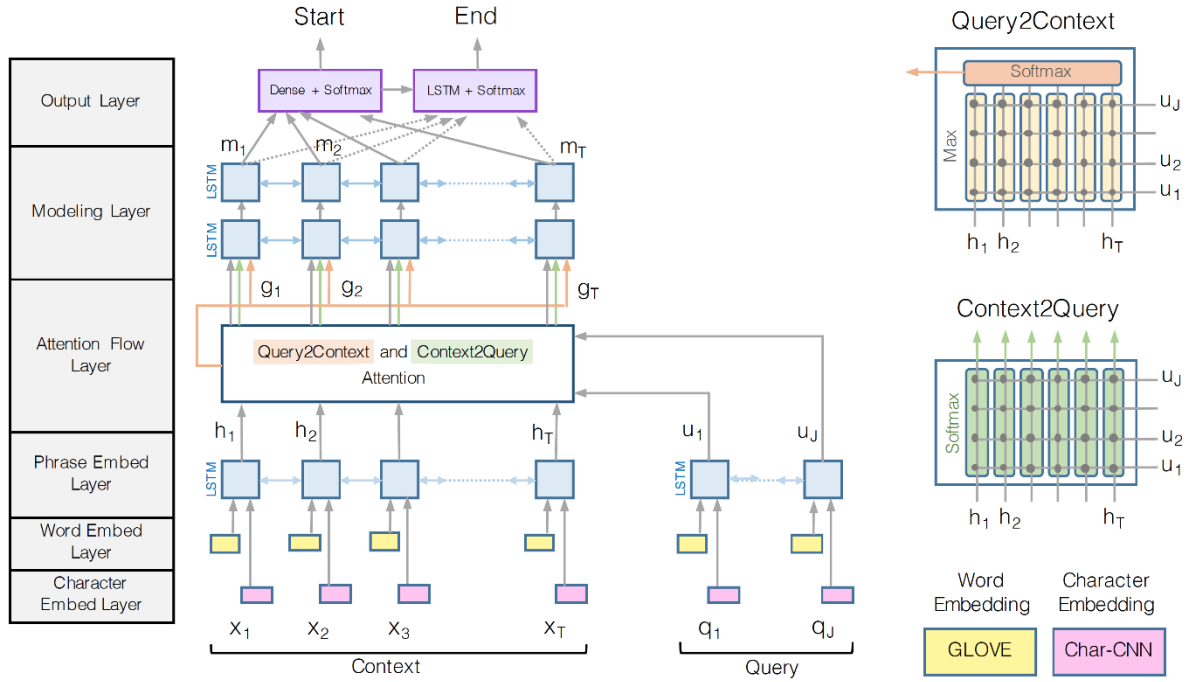
Answer: through contact with Persian traders

How to QA?

QA 시스템이 어떻게 동작하는지 알아보자!

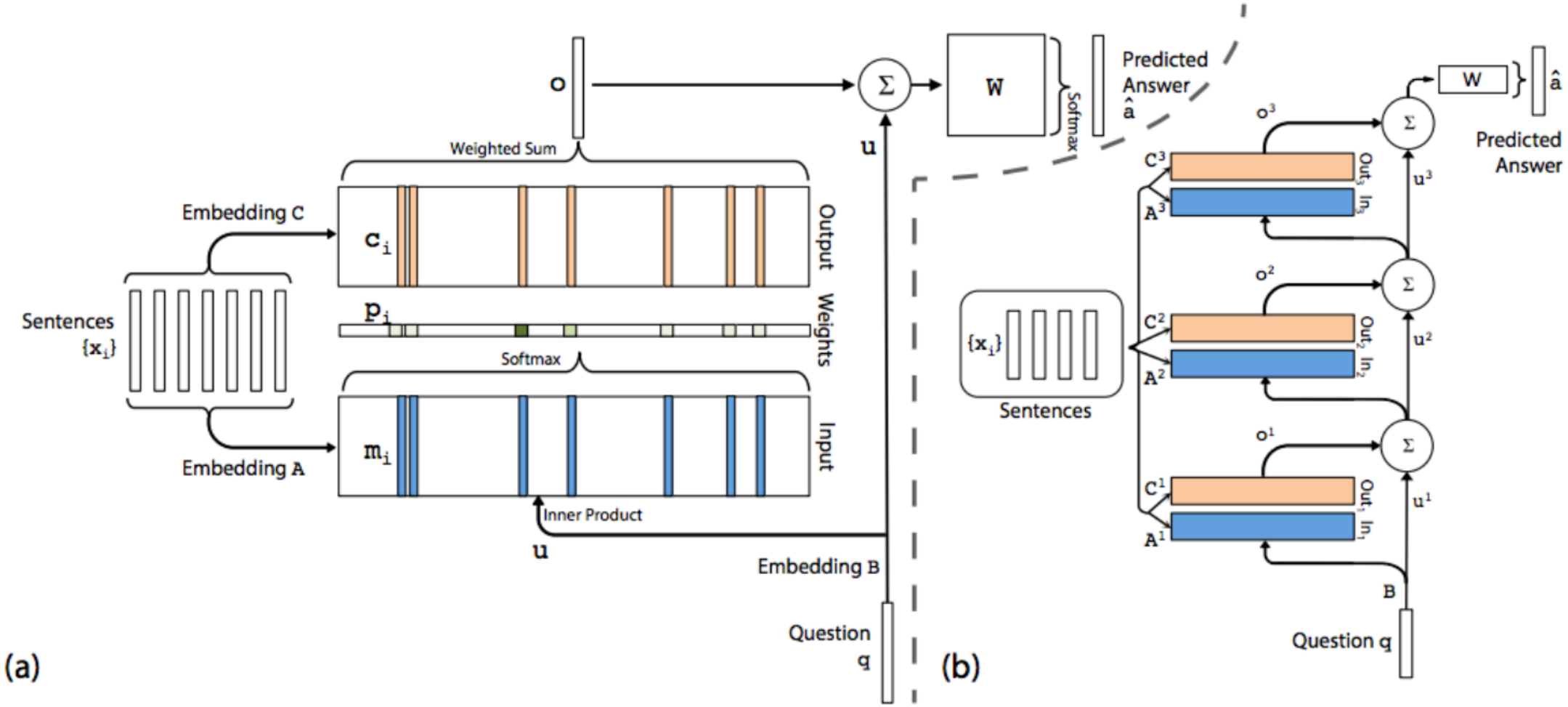


❖ 입력된 질의를 기계가 어떻게 받아들이는지 (자연어 처리)

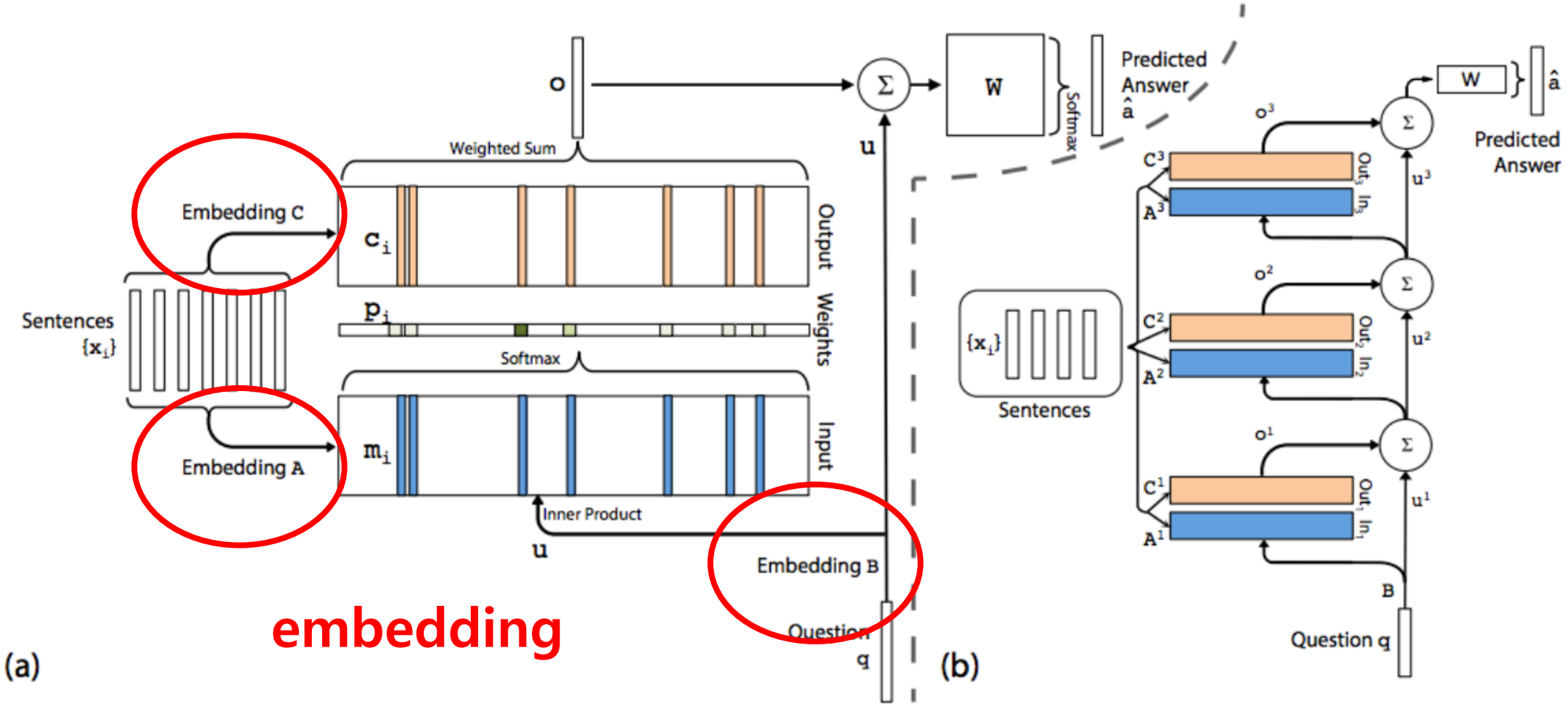


❖ 질문에 대한 답을 제시하기까지 어떠한 과정을 거치는지

MemN2N

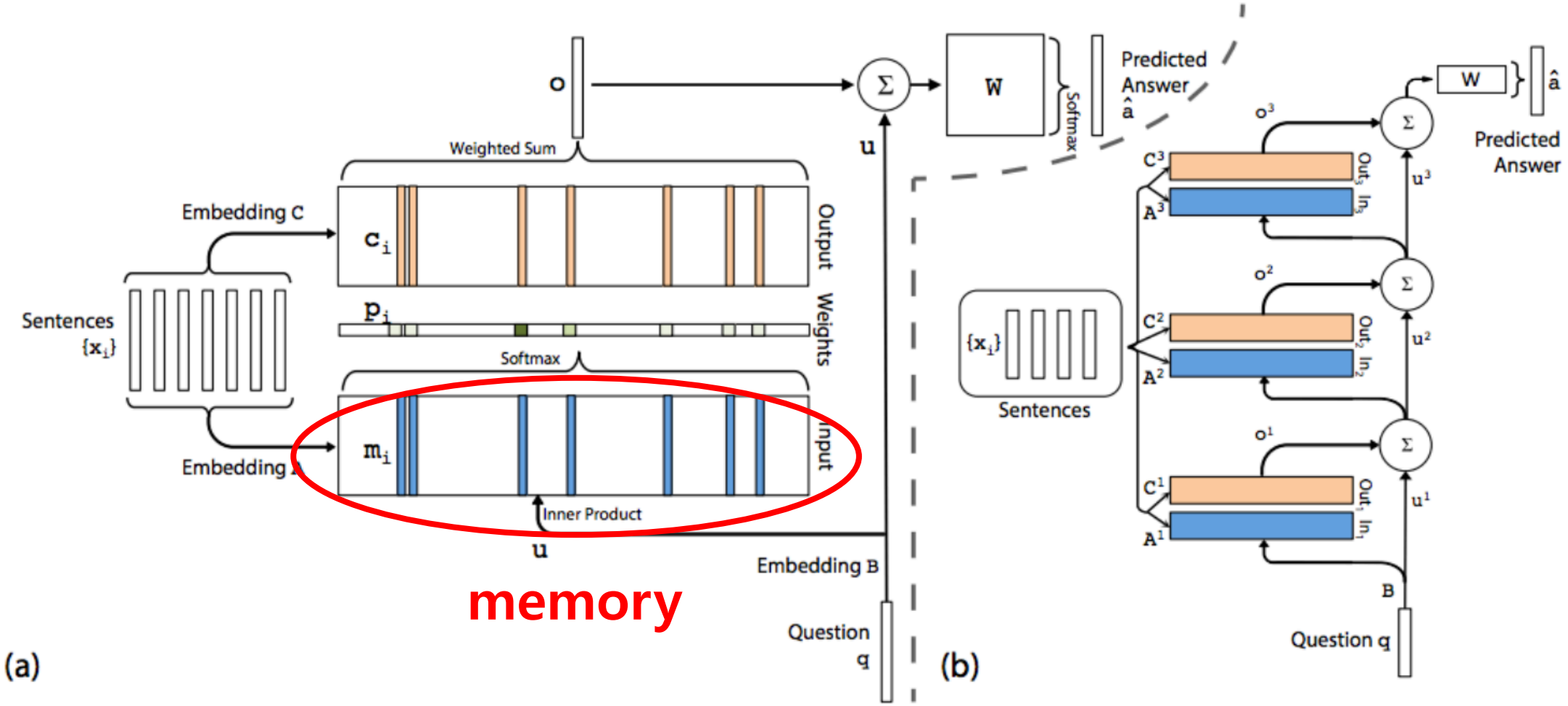

o
Σ
W
Softmax
Predicted Answer
â
Weighted Sum
u
Embedding C
Sentences
{x_i}
c_i
Output
p_i
Weights
Softmax
m_i
Input
Embedding A
Inner Product
u
Embedding B
Question
q
(a)
{x_i}
Sentences
C^3
A^3
C^2
A^2
C^1
A^1
o^3
o^2
o^1
u^3
u^2
u^1
B
Question q
(b)

MemN2N



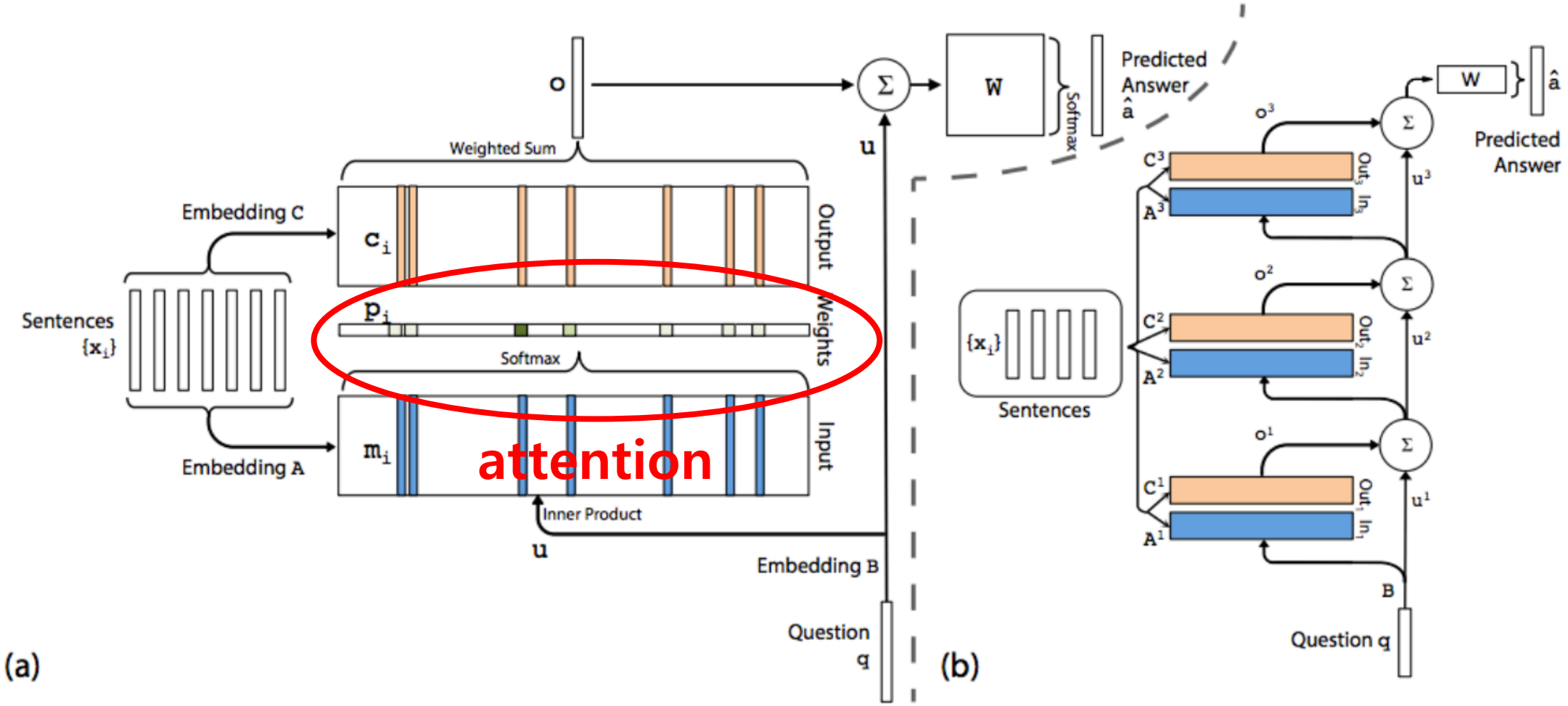
o
Σ
W
Softmax
Predicted Answer
â
u
Weighted Sum
Embedding C
c_i
Output
Sentences
$\{x_i\}$
p_i
Weights
Softmax
Embedding A
m_i
Input
Inner Product
u
Embedding B
Question q
(a)
embedding
(b)
$\{x_i\}$
Sentences
C^3
A^3
C^2
A^2
C^1
A^1
o^3
o^2
o^1
u^3
u^2
u^1
B
Question q
W
â
Predicted Answer

MemN2N



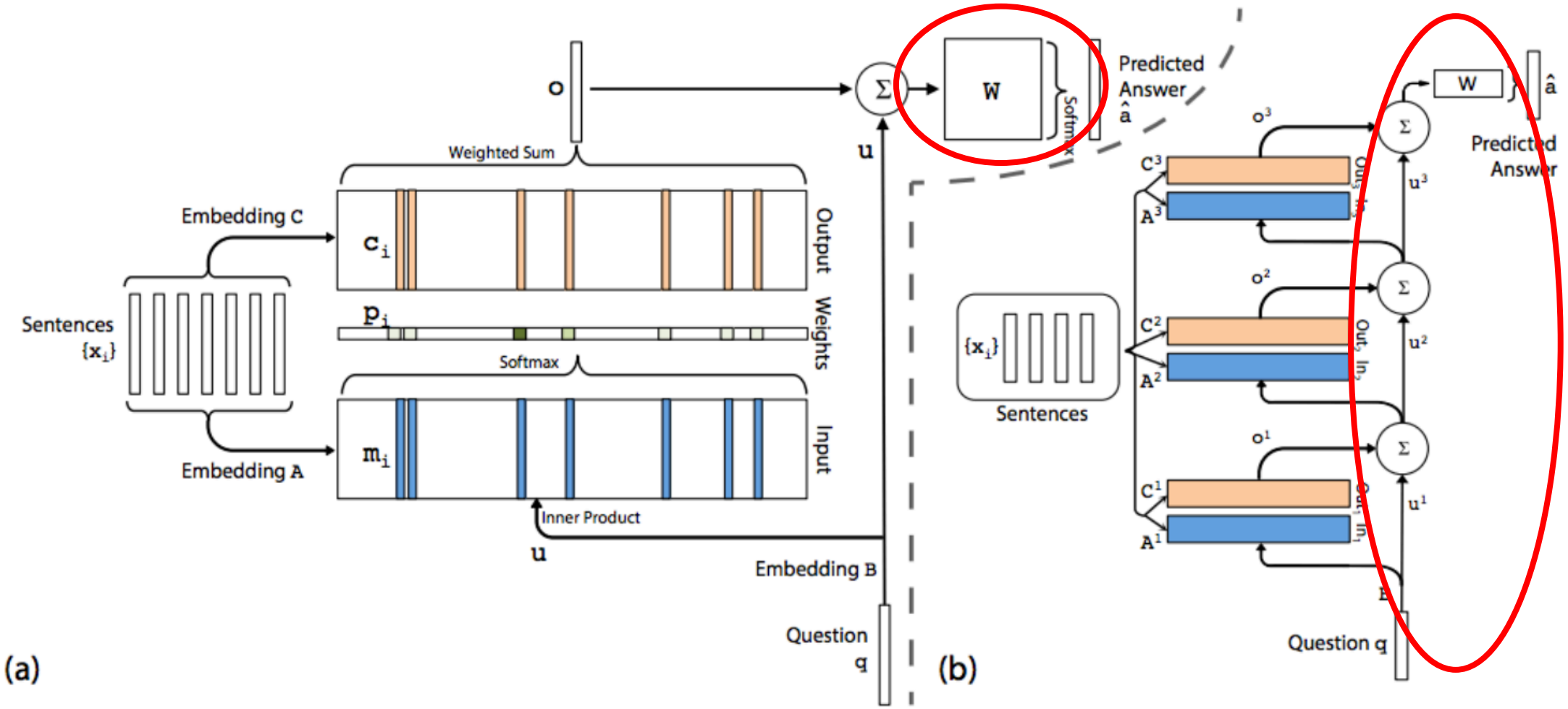
o
Σ
W
Softmax
Predicted Answer
â
u
Weighted Sum
Embedding C
c_i
Output
Sentences
$\{x_i\}$
p_i
Weights
Softmax
m_i
Input
Embedding A
Inner Product
u
Embedding B
memory
Question q
(a)
(b)
o^3
C^3
A^3
Out_3
In_3
u^3
o^2
C^2
A^2
Out_2
In_2
u^2
o^1
C^1
A^1
Out_1
In_1
u^1
B
Question q
Predicted Answer

MemN2N


o
Σ
W
Softmax
Predicted Answer
â
Weighted Sum
u
Embedding C
Sentences
{x_i}
c_i
p_i
m_i
Output
Weights
Input
Softmax
attention
Embedding A
Inner Product
u
Embedding B
Question q
(a)
{x_i}
Sentences
C^3
A^3
C^2
A^2
C^1
A^1
o^3
o^2
o^1
u^3
u^2
u^1
B
Question q
(b)

MemN2N

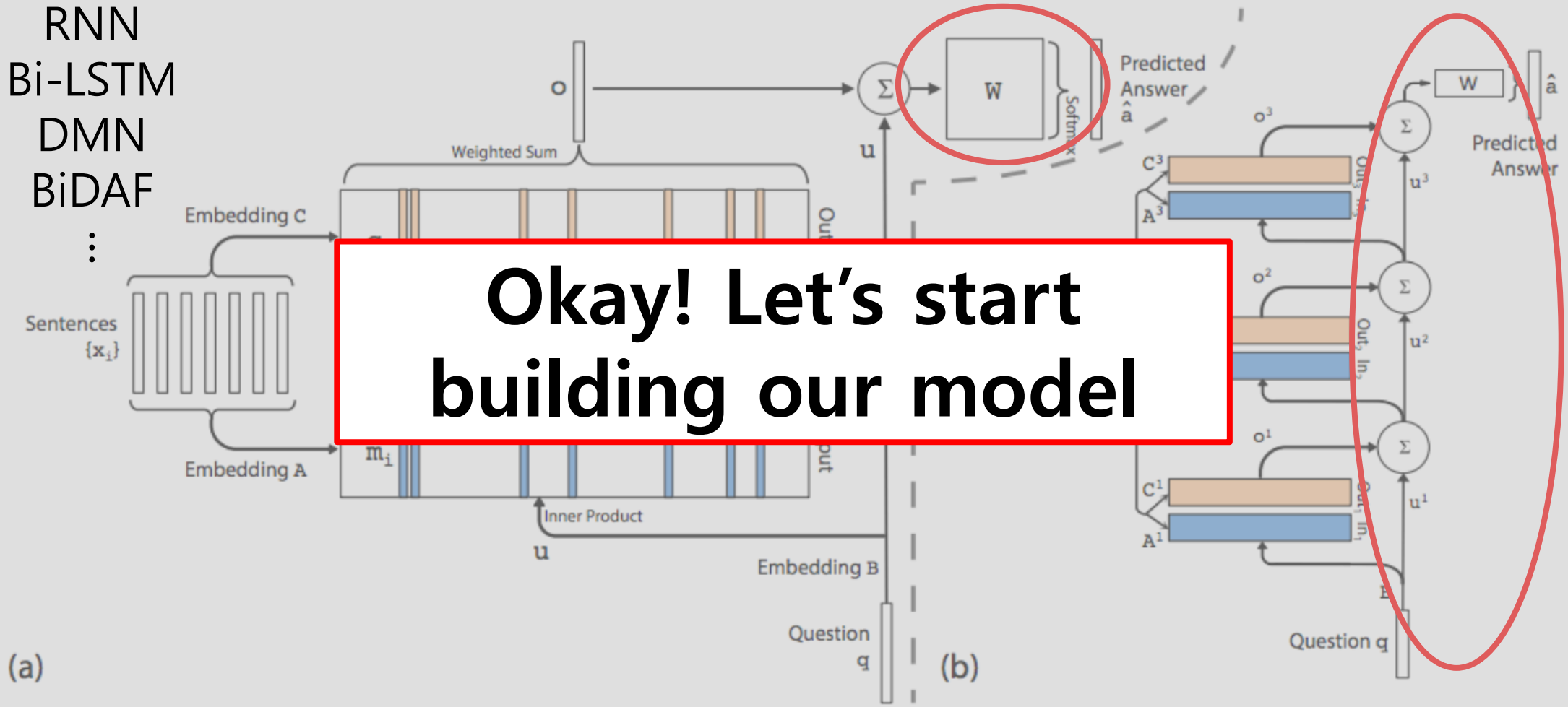
Another layer
Or multi hops



MemN2N

RNN
Bi-LSTM
DMN
BiDAF
⋮

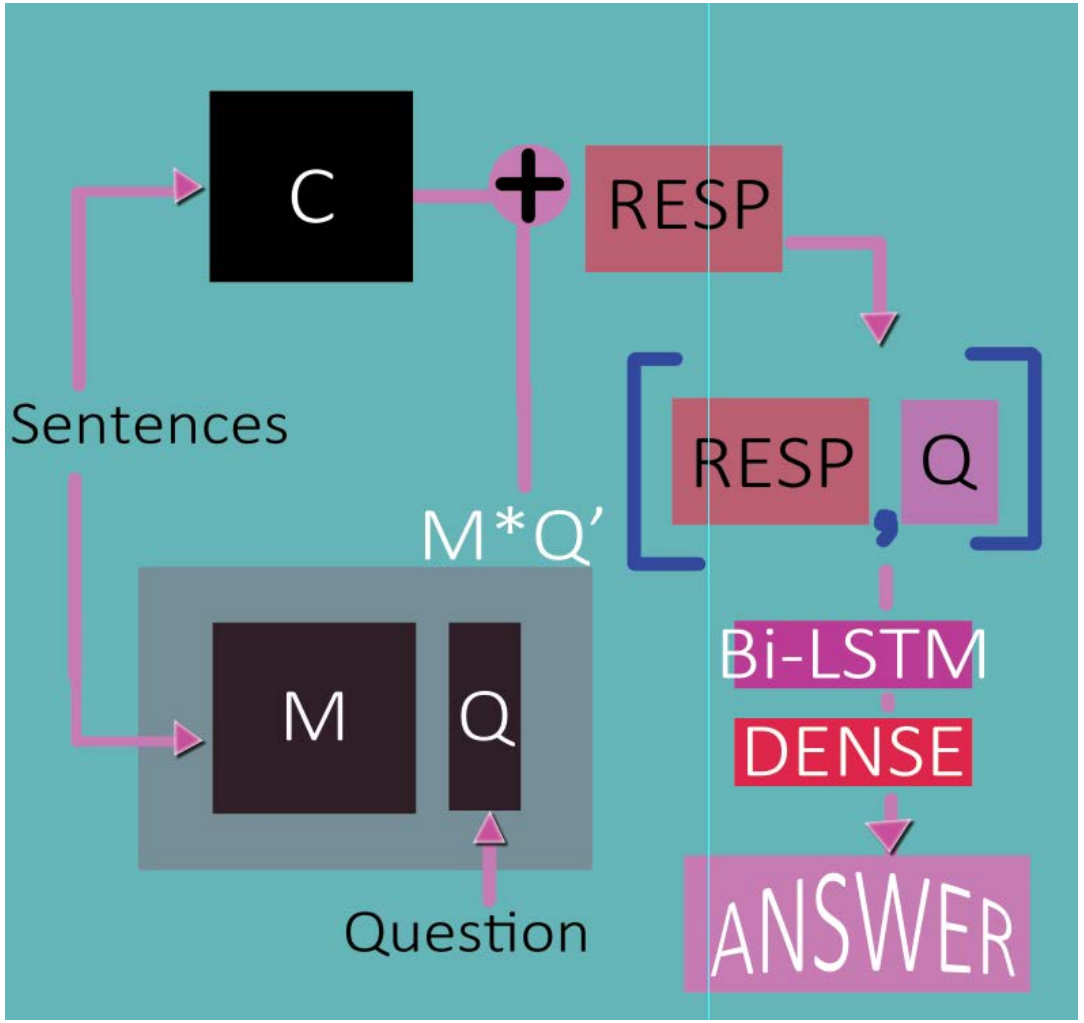
Another layer
Or multi hops



Okay! Let's start building our model

MemN2N – Ours

- C: Word-embedded sentences
- M: Word-embedded sentences
- Q: Word-embedded sentences
- M*Q': 임베딩된 질의와 임베딩된 문장 사이의 Similarity 곱
- RESP: M*Q'와 C의 합
- [RESP, Q]: RESP와 Q의 Concatenation



MemN2N – Ours

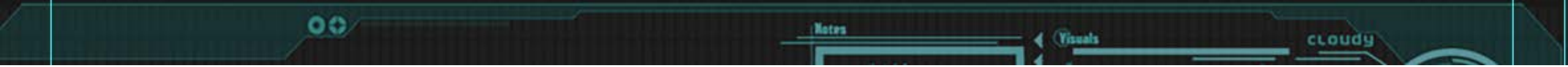
https://arxiv.org/ftp/arxiv/papers/1703/1703.03939.pdf

Task	SOTA	DMN	OURS
1. Single supporting fact	100%	100%	93.9%
2. Two supporting facts	81.2%	39.7%	39.2%
3. Three supporting facts	68.3%	41.5%	26.6%
4. Two argument relations	82.5%	75.5%	77%
5. Three arguments relations	87.1%	50.1%	87.3%
6. Yes/no questions	**98%**	97.7%	84.2%
7. Counting	89.9%	91.4%	83.9%
8. Lists/sets	93.9%	95.2%	75.8%
9. Simple negation	98.5%	99%	64.5%
10. Indefinite knowledge	97.4%	87.3%	89.6%
11. Basic coreference	96.7%	100%	75.2%
12. Conjuction	100%	87%	77.2%
13. Compound coreference	99.5%	96.4%	94.4%
14. Time reasoning	98%	73.1%	43.5%
15. Basic deduction	98.2%	53.9%	59.1%
16. Basic induction	49%	49.5%	51.7%
17. Positional reasoning	57.4%	59.3%	62.3%
18. Size reasoning	90.8%	98.3%	92.1%
19. Path finding	9.4%	9%	9.5%
20. Agent's motivations	100%	97.1%	98.6%
Average accuracy	84.775%	75.05%	69.28%

MemN2N – Ours

- End-to-end
- Simple and light (Single hop)
  - No fine tuning, no complicated guiding (Linear Start, Positional Encoding, Temporal Encoding, Random empty Memory)
- Fast and efficient ~~(if it can be trained)~~
  - avr 400 ~ 600s
- Reasonable performance
  - Some are better than SOTA

MemN2N – Web Demo



DEMO

STORY

Mary went back to the hallway .
Sandra went back to the garden .
Mary went to the office .
Daniel travelled to the garden .
Mary went to the bedroom .
Sandra went back to the kitchen .

QUESTION

Where is Daniel ?

ANSWER

Answer = 'garden'
Confidence score = 93.91%
Correct!

Predict | New story

MemN2N – Ours

1. ~~**Build a working QA system from a scratch**~~
   - ~~Problem se~~
   - ~~Understand~~ ~~hm~~
   - ~~Knowing the current state of the art~~
2. **Attack more advanced application or problem set**
   - SQuAD
   - Knowledge based + DL QA system

Okay, It works!
Let's move on.

How to QA? – Problem Setup

Facebook: bAbI

Two supporting facts

story

not really...

BUSTED

1 Mary got the milk there.
2 John moved to the [illegible]
3 [illegible]
4 [illegible] to the hallway

5 Where is the milk? question

answer

hallway | 1 4 supporting line

SQuAD

The Stanford Question Answering Dataset

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders

SQuAD

위키피디아 500개 이상의 포스팅으로부터 만들어진
100,000개 이상의 질문과 정답 세트

지문이 주어지고, 그를 바탕으로 답할 수 있는 질문과 답이 주어짐

위키피디아를 바탕으로 여러 사람들 (crowd workers)이 직접 질문과 답을 작성

이로 인해 질문의 내용을 확장 이해해야 답을 할 수 있는 경우, 문장의 구조가 바뀌어서 문맥을 추론해야 답을 할 수 있는 경우 등의 고난이도 케이스가 포함되어 있음

SQuAD

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

Which governing bodies have veto power?

❖ Governing bodies가 The Europenan Parliament and the Council of the EU를 지칭함을 알아야 답할 수 있는 경우

Passage Segment

...The Rankine cycle is sometimes referred to as a practical Carnot cycle...

Question

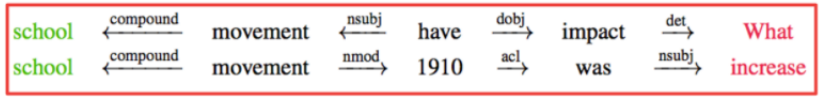
What is the Rankine cycle sometimes called?

❖ 지문의 referred to as 가 질문의 called 가 동일한 의미임을 알아야 하는 경우

Syntactic Variation

...During the mass high school education movement from 1910 – 1940 , there was an increase in skilled workers...

What impact did the high school education movement have on the presence of skilled workers?

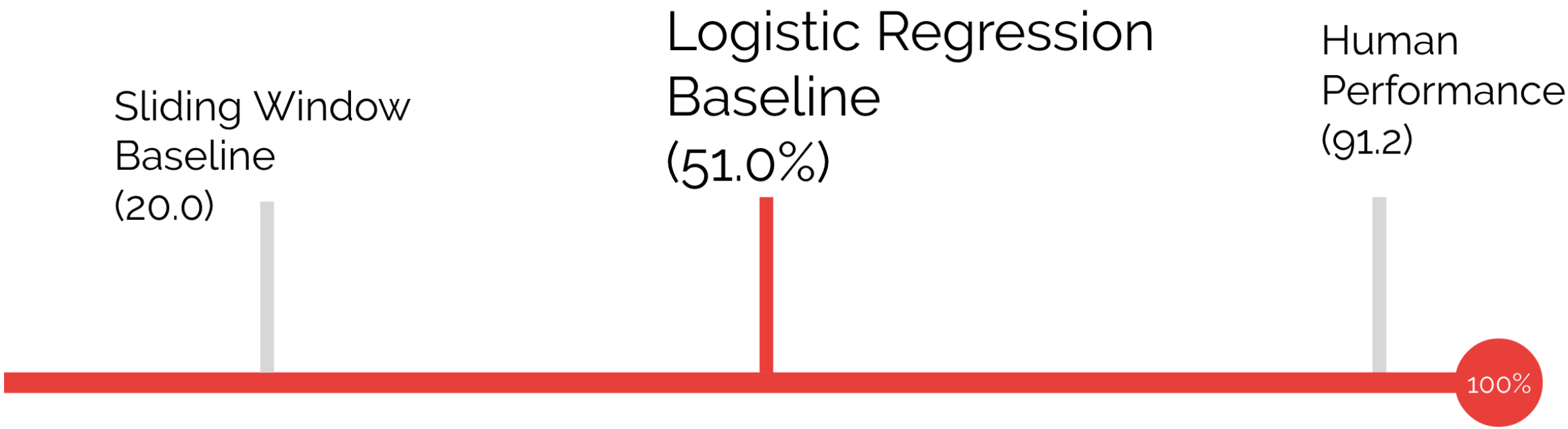


❖ 질문의 문장 구조가 지문에서 정답을 포함한 문장의 구조가 달라진 경우

SQuAD: Data preprocessing

- Huge txt data
- Human errors in the dataset (wrong answer locations)
- Lots of special characters
  - Latin char, 0-9
- Stopwords (ablation)
- Stemming (ablation)

SQuAD: SOTA?

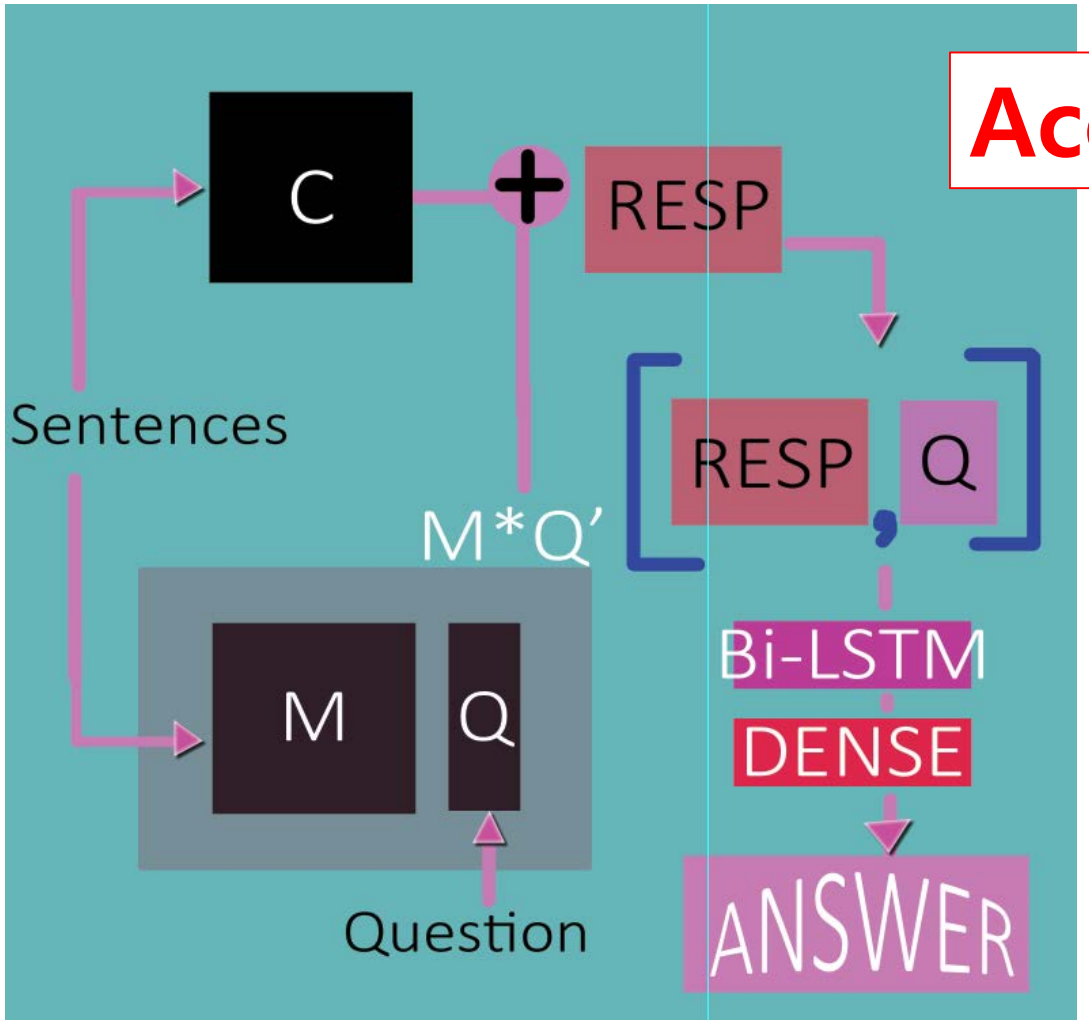


Metric

EM: Exact Match

F1: Average overlap between prediction and ground truth

MemN2N – Ours



Acc: ~8%

Why?: Problem shooting

Path finding

story

1 The garden is west of the bathroom.
2 The bedroom is north of the hallway.
3 The office is south of the hallway.
4 The bathroom is north of the bedroom.
5 The kitchen is east of the bedroom.

question

6 How do you go from the bathroom to the hallway?

s,s

answer

4 2

supporting line

Acc: ~9.5%

Why?: Problem shooting

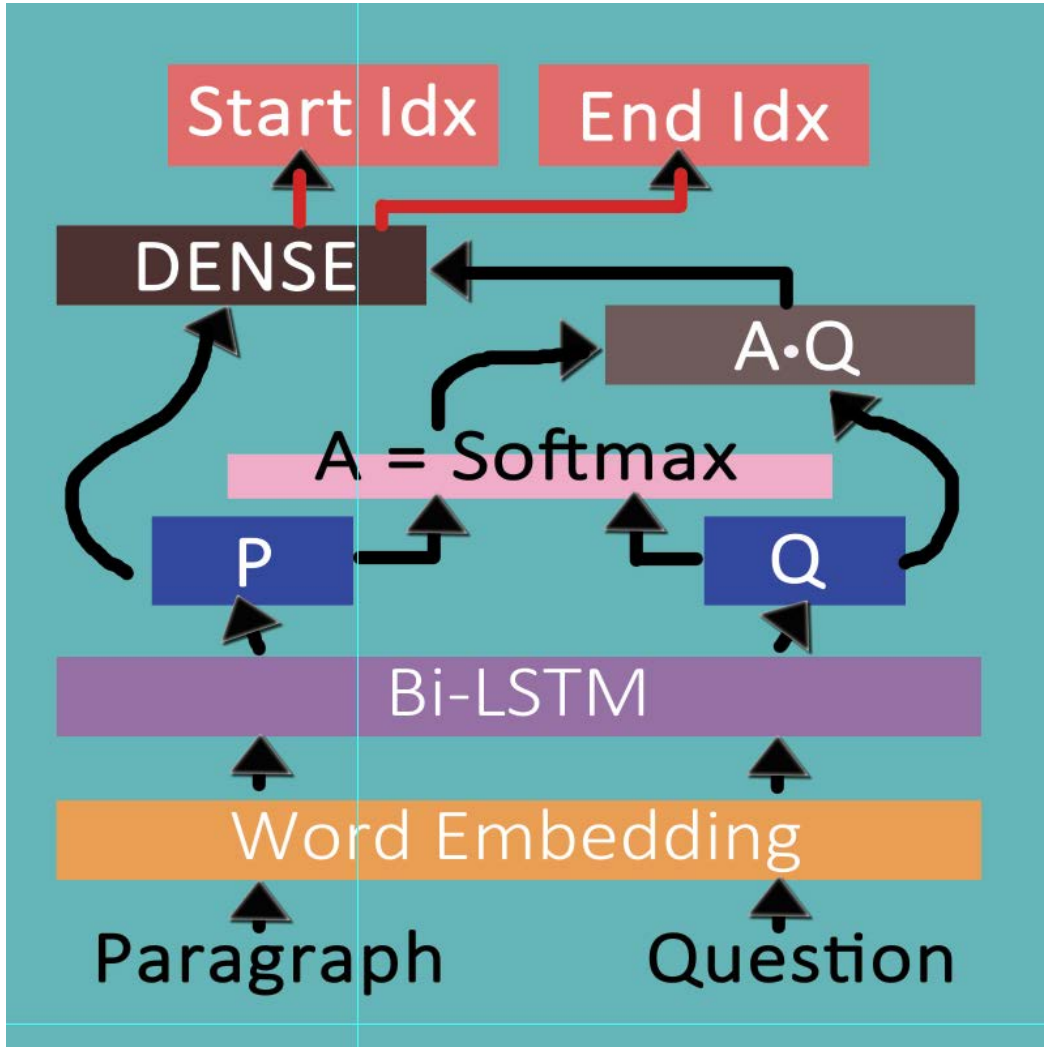
1 super bowl 50 was an american football game to determine the champion of the national football league nfl for the 2015 season the american football conference afc champion denver broncos defeated the national football conference nfc champion carolina panthers 24 10 to earn their third super bowl title 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

which nfl team represented the afc at super bowl 50

denver broncos 28 29

MemN2N – Ours

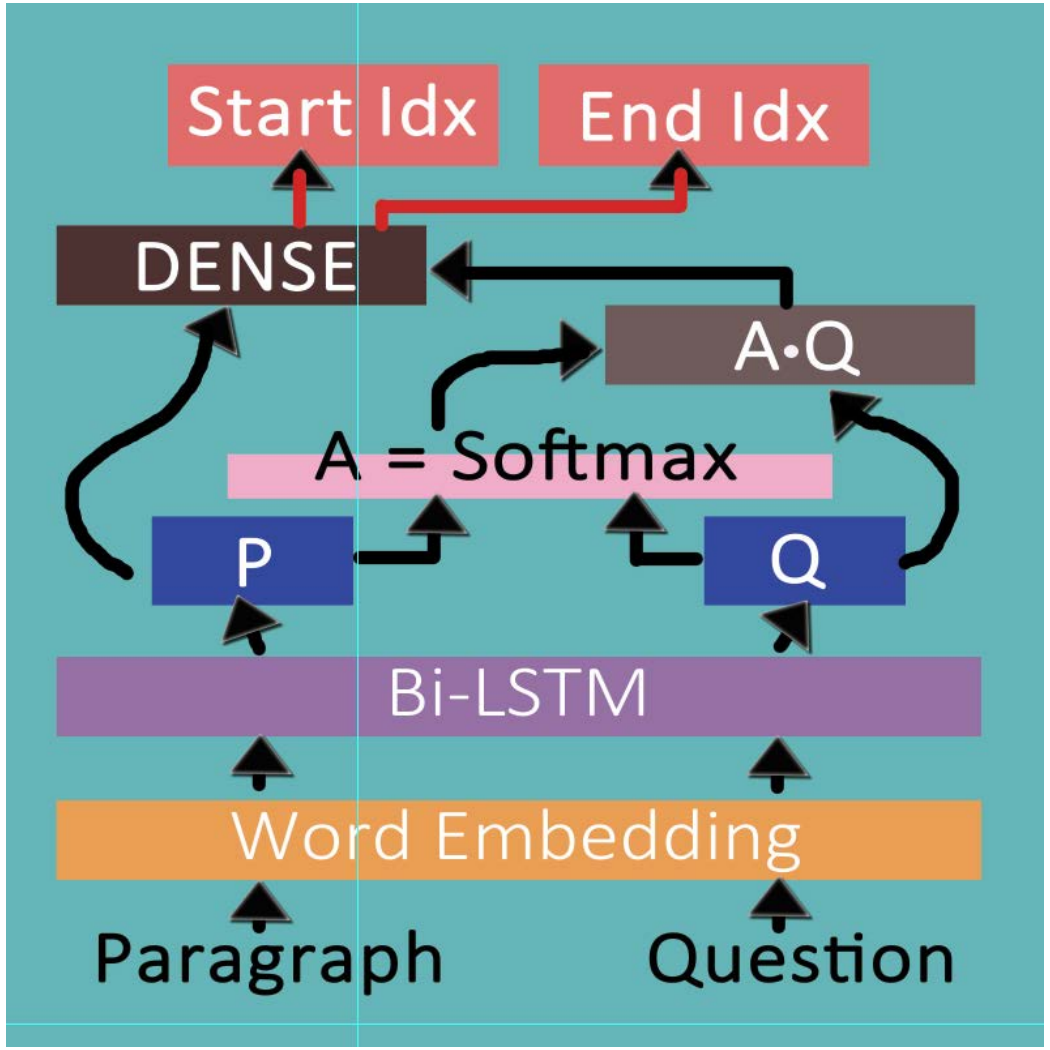
- P: 제시된 지문 W2V-embedding, Bi-LSTM Output
- Q: 질의에 대해 W2V-embedding, Bi-LSTM Outpuy
- A: Attention
- A Q: A와 Q 사이의 내적
- Start Idx: 지문 내 정답 시작 위치
- End Idx: 지문 내 정답 끝 위치



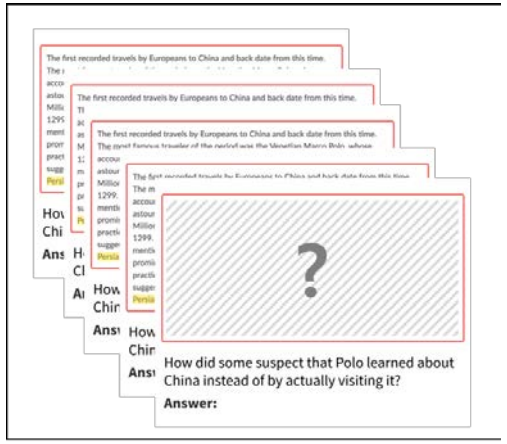
MemN2N – Ours

STILL ON GOING

- P: 제시된 지문 W2V-embedding, Bi-LSTM Output
- Q: 질의에 대해 W2V-embedding, Bi-LSTM Outpuy
- A: Attention
- A Q: A와 Q 사이의 내적
- Start Idx: 지문 내 정답 시작 위치
- End Idx: 지문 내 정답 끝 위치



Future work? Remaining Idea



$\{P, Q\}: A?$

Parse Q and find its keyword
Search Wiki &
Crawl the abstract passage

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

P'

$\{P', Q\}$ ➡ MODEL ➡ *Answer*

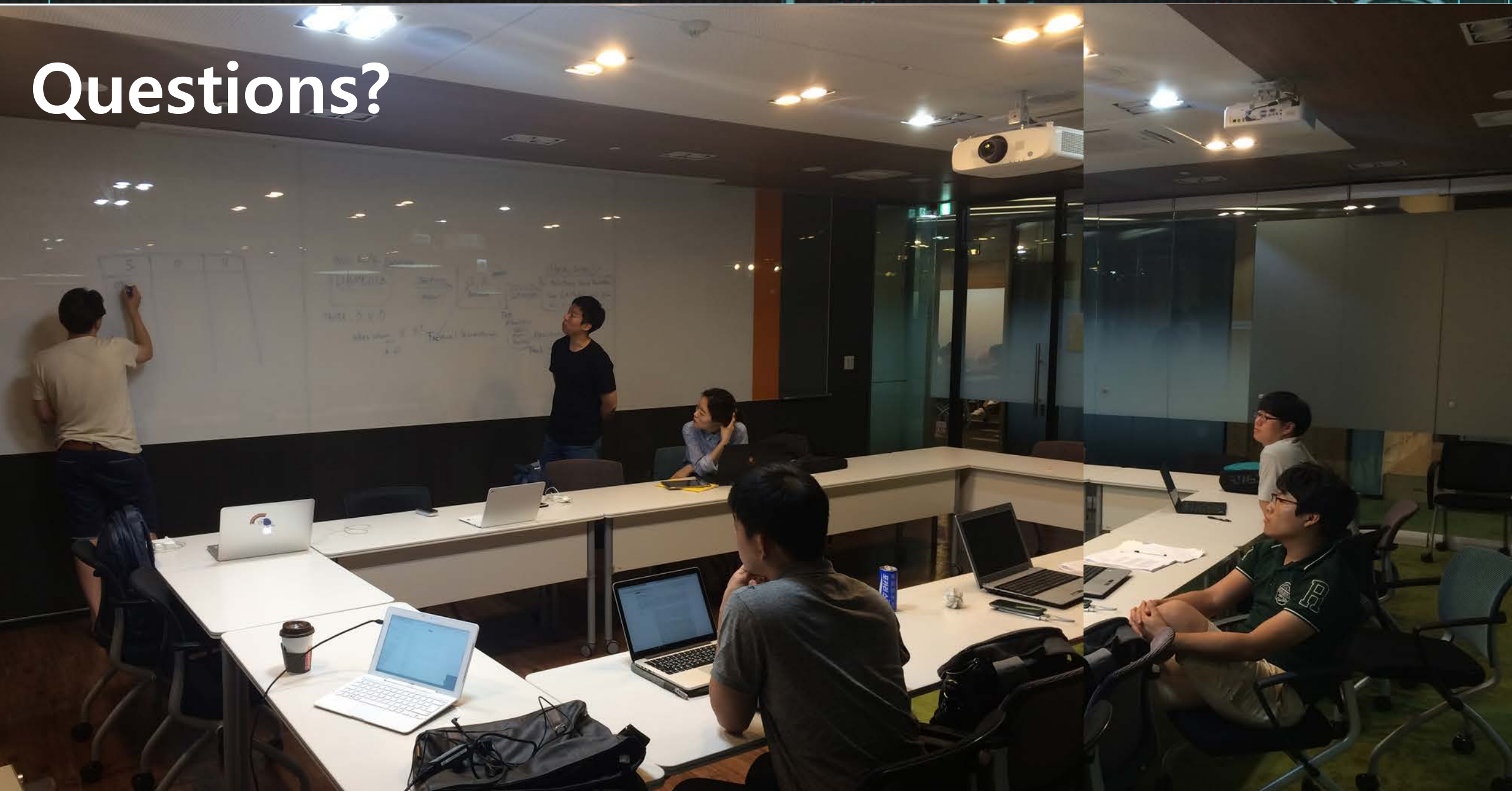
Define a new problem

- **Do not have to compete with the existing tasks!** $+ \alpha$ **!**
- **Creativity** $+ \alpha$ **!**
- **Plug & Play** $+ \alpha$ **!**
- **More realistic setup** $+ \alpha$ **!**
- **Better performance** $+ \alpha$ **!**

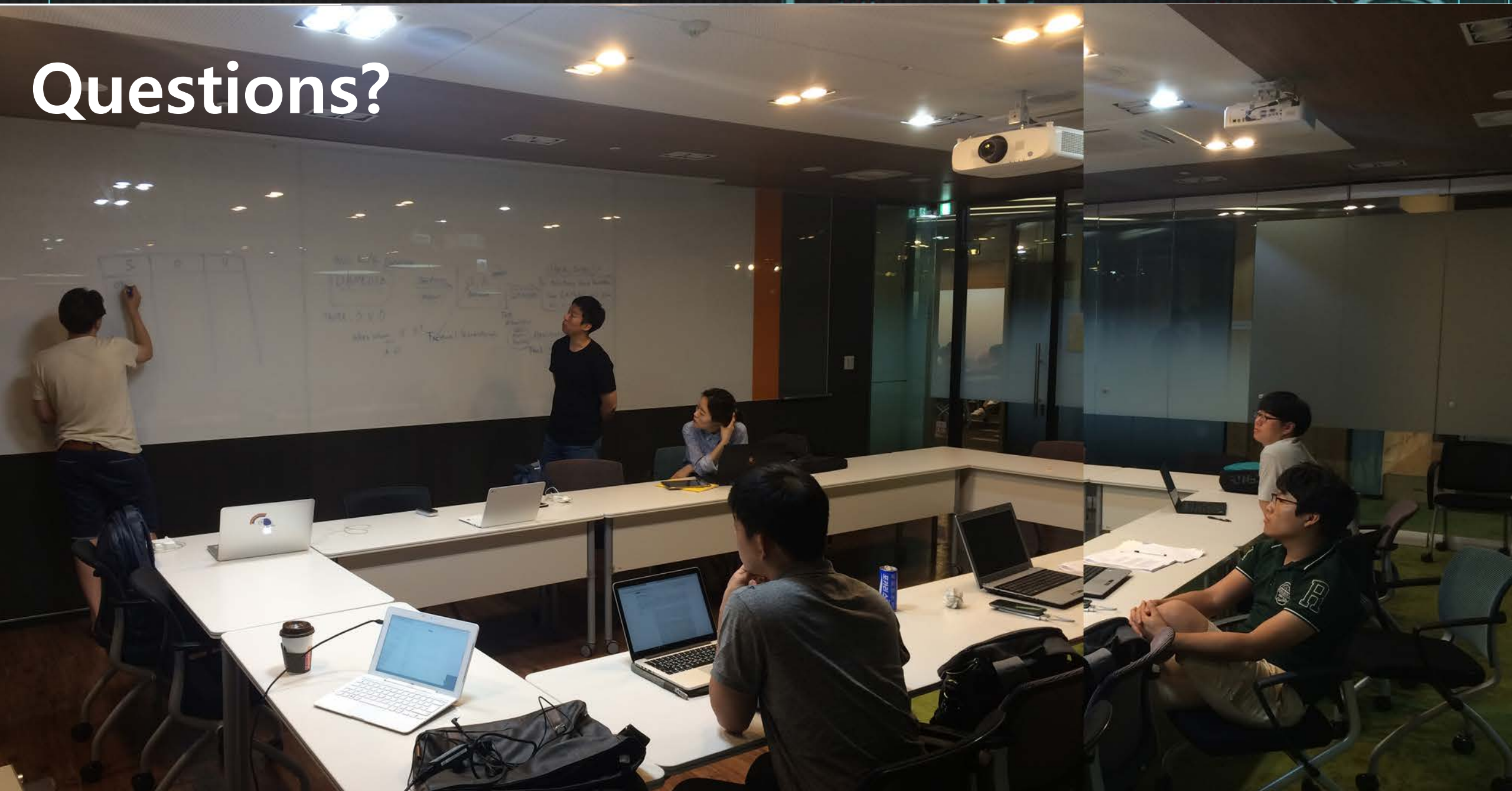
Summary

1. Managed to achieve a goal in short term
2. Build a working QA system from a scratch
3. Played with whole system from data preprocessing to final working product

Questions?



Questions?



SQuAD: SOTA?

Leaderboard

Since the release of our dataset, the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test and development sets of v1.1. Will your model outperform humans on the QA task?

Rank	Model	EM	F1
1 Aug 16, 2017	DCN+ (ensemble) *Salesforce Research*	78.706	85.619
2 Jul 25, 2017	Interactive AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	77.845	85.297
3 Aug 21, 2017	Reinforced Mnemonic Reader (ensemble) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798	77.678	84.888