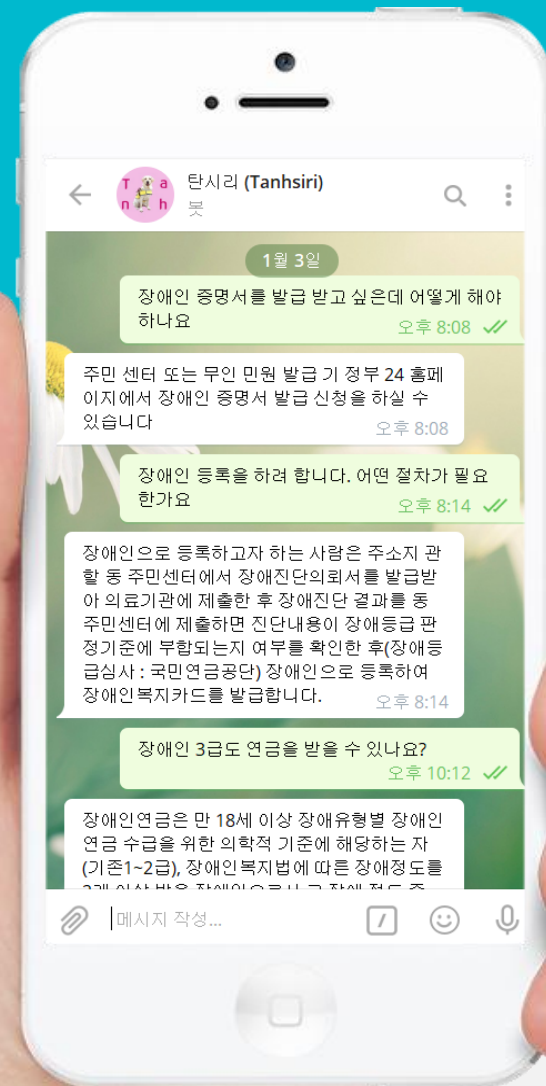


장애인의 용이한 정보접근을 위한 ChatBot

Tanh-siri

이준걸 임진혁 권혜민

김효은 배유나 신윤종 이유진



목차

01

문제 정의

02

데이터

03

모델링

04

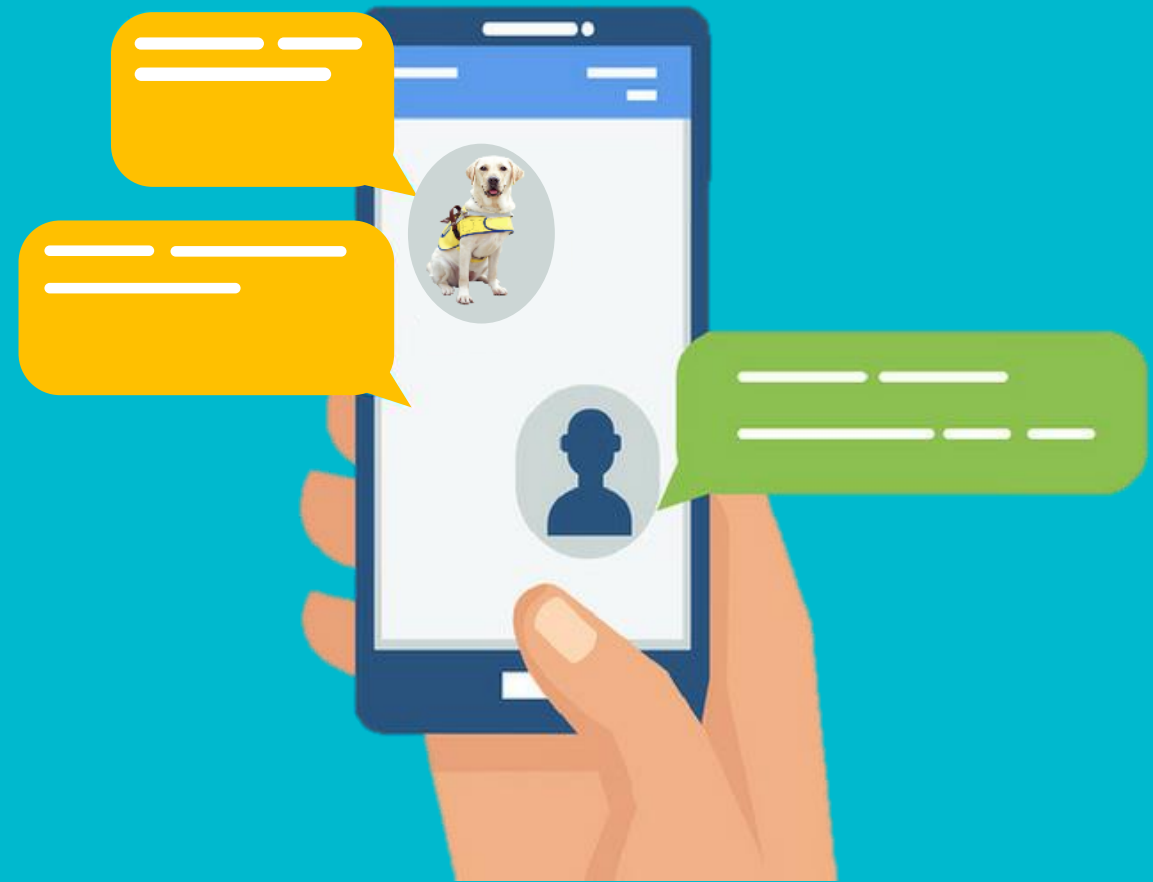
시연

05

기대효과 및 한계점

Part1.

문제 정의



문제 정의

장애인의 낮은 정보 접근성

[메이블뉴스 - 몰라서 못 받는 연말정산 '중증환자 장애인공제'](#)

<http://www.ablenews.co.kr/News/Include/NewsContentInc.aspx?NewsCode=00,...> ▼

병원에서 **장애인**증명서 발급에 대해 모르고 있어 발급받지 못 했다. 이후 2011년에... 현재 갑상선이나 각종 암 등 **중증** 질환을 치료받은 경우, 세법상 **장애인**으로 분류돼 소득**공제**를 받을 수 있으나, 현재 **연말정산** 간소화시스템에서 조화가 되지 않고...



[장애인 근로자 올린 복지부, 1만5,916명 연차수당 미지급](#) 장애인신문 | 5일 전 |

이에 따라 **장애인** 일자리사업에 참여한 1년 계약자에게는 11개월간 발생한 11일의 연차휴가와 1년 시점에서... 감안하여 **법** 적용 여부에 대해 좀 더 면밀히 검토한 후 처리 방안을 결정한다는 입장이다. 현행법상 1년 동안...



[부산지역 장애인, "30대 중차한 두리발, 현실은 여전히 불편"](#)

노컷뉴스 | 5시간 전 | 네이버뉴스 |

사상구**장애인**자립센터는 "부산시의 교통약자 이동 편의 정책 대대적인 홍보에도 불구하고, 두리발 배차시간이 전보다 더 길어져 **장애인** 등 이용자들이 여전히 불편을 겪고 있다"고 24일 밝혔다. 부산시는 지난 7월...

문제 정의

기존 Q&A 플랫폼의 느린 답변 대기 시간

- 네이버 지식인

'장애인'에 대해 답변을 기다리는 질문입니다. [+ 관심 키워드 추가](#)

제목	분야	답변	등록일
올해 장애인연금 인상 되나요.?	장애인복지	0	10시간 전
장애인연금 언제 인상 되나요.?	장애인복지	0	10시간 전
장애인연금 인상 안됐나요.?	장애인복지	0	10시간 전
50 장애인활동지사	장애인복지	1	10시간 전
100 자동차 이전 및 장애인주차표찰 받부	자동차세	0	12시간 전
100 시각장애인 외국 방문시 출입국심사 관련...	외국법	0	20시간 전
50 장애등급 폐지되면 경증 정신장애인도...	장애인복지	1	21시간 전
차상위장애인응급실진료비 두통이 너무심	응급의학과	0	22시간 전
100 지적장애인인 제 친구랑 아는 형	특수교육	0	23시간 전
저희반 장애인 때문에 너무 힘들어요	장애인복지	0	23시간 전
장애인도 기상캐스터에 도전할 수...	직업, 취업	0	2020.01.02.
90 시청에서 장애인일자리하면서 부업으로	마산시	0	2020.01.02.
50 장애인 등록되었는지 궁금합니다.	장애인복지	1	2020.01.02.

- 국민 신문고

민원 처리기간은 어떻게 되나요?

일반적으로 접수일로부터 7일에서 14일 이내에 처리되나 토요일·공휴일은 제외되며, 민원의 성격이나 해당 민원을 처리하는 기관의 규정에 따라 일부 달라질 수 있습니다. 아울러 부득이한 사유로 처리기간 내에 민원을 처리하기 곤란하다고 인정되는 경우에는 해당 처리기간의 범위에서 처리기간을 연장할 수 있습니다.



① 민원 처리기간표

민원유형	처리기간
법령에 관하여 설명이나 해석을 요구하는 질의민원	14일 이내
제도·절차 등 법령 외에 사항에 관하여 설명이나 해석을 요구하는 질의민원	7일 이내
행정제도 및 운영의 개선을 요구하는 민원	14일 이내
고충민원(행정기관 등의 위법·부당하거나 소극적인 처분 및 불합리한 행정제도로 인하여 국민의 권리를 침해하거나 불편, 부담을 주는 사항에 관한 민원)	7일 이내

문제 정의

장애인에게 필요한 정보를 제공하는 챗봇 Tanh-siri



T

a

n

h



Part2. 데이터



데이터

Data collection



■ 장애인 자동차 표지 발급

가족 중에 장애인이 있는데 장애인 자동차 표지를 발급 받을 수 있나요?

906개



Q 국민연금 장애인연금 질문 2019.11.29.

일반인이었다가 뇌병변 3급 장애인이 되었습니다. 이경우 국민연금과 장애인연금을 동시에 받을 수 있나요?
A 우선적으로 뇌병변 3급이라 하시는 부분은 국가장애인 장애인복지법에 의한... 노령연금과 장애인연금은 각각 수령이 가능합니다. 장애인연금은 장애인복지법에...

지식파트너 답변 국민연금공단

75254개



[장애인복지] 장애인활동지원 수급자격의 유효기간이 있나요?

장애인활동지원 수급자격의 유효기간이 있나요?

A. 장애인활동지원 수급자격의 유효기간은 3년이며, 3년마다 기간 내 신청을 통해 수급자격 갱신을 해야합니다. 또한, 수급자가 65세가 되는 경우 그 해당 월의 다음 월 말일까지 수급자격을 인정합니다. 다만, 수급자격 유효기간...

답변 1 | 추천 0 | 2019.11.04

8312개

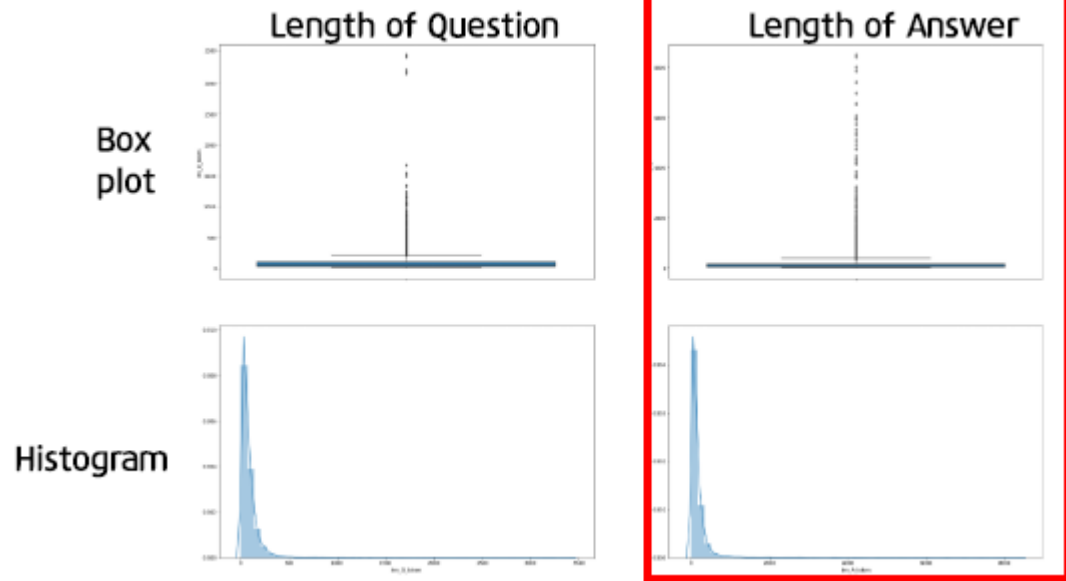


약 90000개 의 Data 확보

데이터

Data problem

- 매우 긴 Answer의 답



- 지나치게 개인적인 사연 & 잘못된 맞춤법

“

제가 중고자동차 하나구입하려고하는데요..제가 가져가야할서류는 뭐죠????그리구여.. 꼭받아야할 서류좀 갈쳐주시구요...또 좋은 중고차 고르는법좀 갈쳐주세여.. 제가 초보인지라.. 글구여 가스차 이짜나여.. 가스차 를 구입하려면 어떻게해야하나는지 좀 갈쳐주세여..장애인증 이써야 하나여???

”

데이터

Data preprocessing

- Question

“제가 중고자동차 하나 구입하려고하는데요..제가 가져가야할서류는 뭐죠????그리구여.. 꼭받아야할 서류좀 알려주시구요...또 좋은 중고차 고르는법좀 알려주세요.. 제가 초보인지라..글구여 가스차 이짜나여.. 가스차 를 구입하려면 어떻게해야하나는지 좀 알려주세요..장애인증 이써야 하나여???”

개인 사연을 일반화 + 맞춤법 정정

“중고자동차 구입할 때 장애인 혜택 받으려면 가져가야 할 서류는 무엇인가요?”

Data preprocessing

- Answer

가족내 세대원중에 한명도 집을 소유한 사람이 없으면, 즉 세대원 모두 무주택이면 장애인 특별공급 대상이 됩니다. 신청은 장애인 명이나 장애인과 같이 거주하는 세대원 명의로 청약신청해도 됩니다. 단 요즘 요건이 강화되어 서울이나 수도권 지역은 세대주만 청약신청이 가능한 곳이 있습니다. 지원대상 무주택 세대 구성원으로서 장애인 등록증이 교부된 자 국민주택 등의 주택을 건설하여 공급하는 경우에 입주자 모집공고일 현재 무주택 세대 구성원으로서 「장애인복지법」 제32조에 따라 장애인 등록증이 교부된 자 등에게 관련 기관의 장이 정하는 우선순위 기준에 따라 한차례에 한정하여 1세대 1주택의 기준으로 그 건설량의 10퍼센트의 범위에서 특별공급함

<https://www.gov.kr/portal/service/serviceInfo/B55201300008>

Data preprocessing

- Answer

~~가족내 세대원중에 한명도 집을 소유한 사람이 없으면, 즉 세대원 모두 무주택이면 장애인 특별공급 대상이 됩니다. 신청은 장애인 명이나 장애인과 같이 거주하는 세대원 명의로 청약신청해도 됩니다. 단 요즘 요건이 강화되어 서울이나 수도권 지역은 세대주만 청약신청이 가능한 곳이 있습니다. 지원대상 무주택 세대 구성원으로서 장애인 등록증이 교부된 자 국민주택 등의 주택을 건설하여 공급하는 경우에 입주자 모집공고일 현재 무주택 세대 구성원으로서 「장애인복지법」 제32조에 따라 장애인 등록증이 교부된 자 등에게 관련 기관의 장이 정하는 우선순위 기준에 따라 한차례에 한정하여 1세대 1주택의 기준으로 그 건설량의 10퍼센트의 범위에서 특별공급함~~

<https://www.gov.kr/portal/service/serviceInfo/B55201300008>

질문에 관련 없는 내용 삭제 + 신뢰성 확보되지 않는 답변 삭제 + 부가적인 링크 삭제 후 약 15000개 데이터 셋 구성

Part3. 모델링



Model

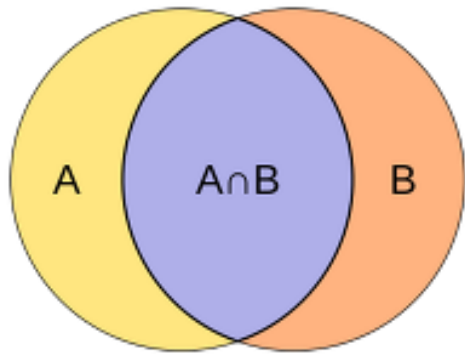
Approach

- 1. Similarity : 질문과 유사도가 가장 높은 질문을 찾아 그 질문의 답변을 챗봇 답변으로 제시
- 2. DL Generation Model : 질문과 문장을 생성 모델에 학습시켜서 질문에 대한 답변 자체를 학습하여 챗봇 답변으로 제시
- 3. Q-A Model : 질문에 대한 답변을 Q-A로 학습시켜 챗봇 답변으로 제시
- 4. Ensemble : 위의 방법들을 적절히 혼합하여 사용

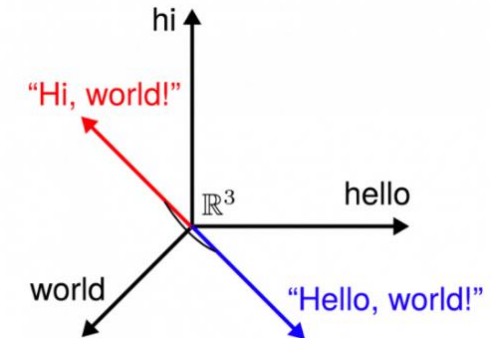
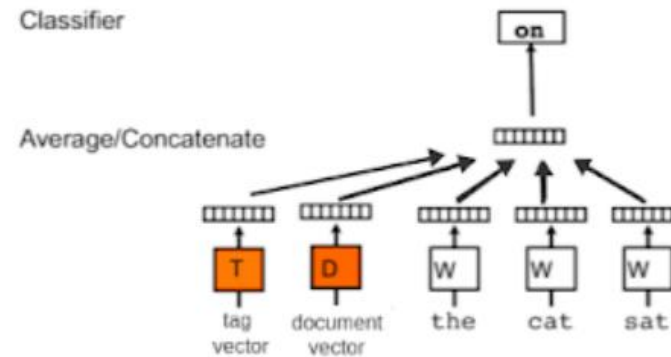
Model

Approach

- 1. Similarity : 질문과 유사도가 가장 높은 질문을 찾아 그 질문의 답변을 챗봇 답변으로 제시



① Jaccard Similarity



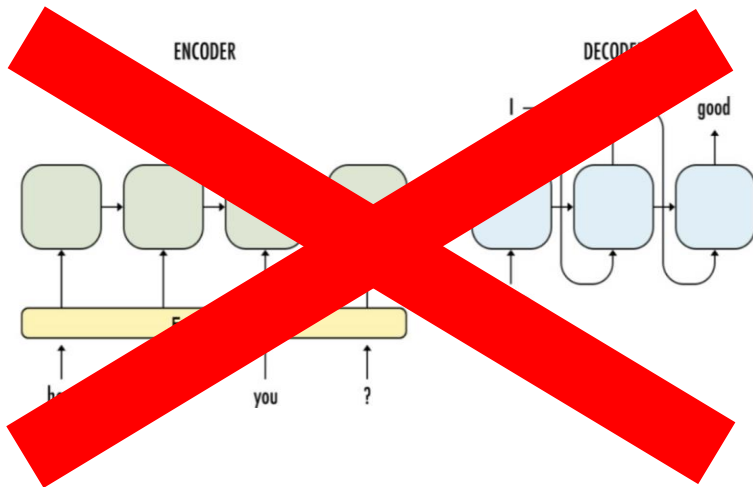
② Embedding 후 Cos similarity

③ Ensemble Method

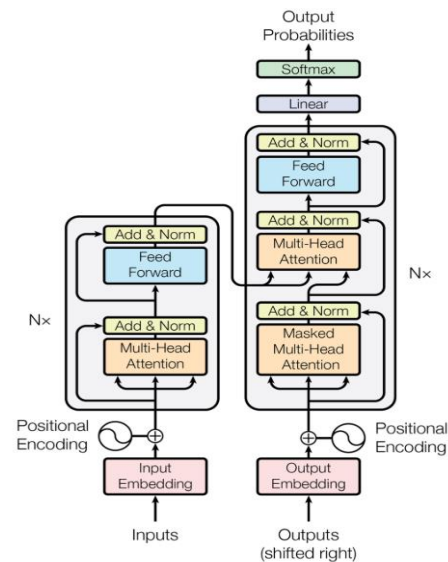
Model

Approach

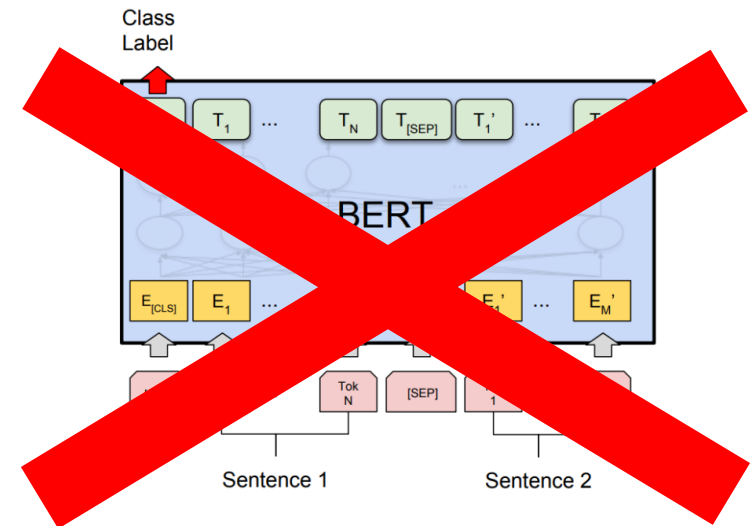
- 2. DL Generation Model : 질문과 문장을 생성 모델에 학습시켜서 질문에 대한 답변 자체를 학습하여 챗봇 답변으로 제시



① Seq2Seq



② Transformer

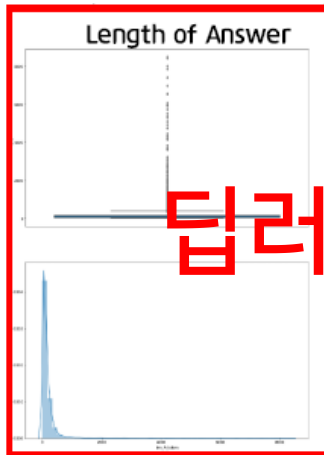
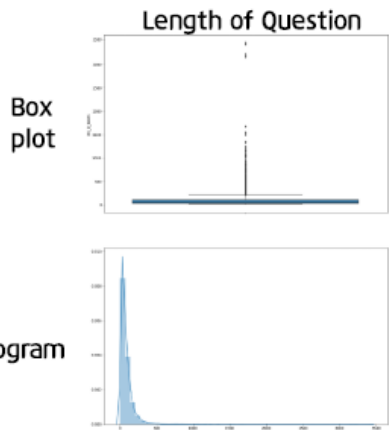


③ BERT

Model

Approach

- 2 DL Generation Model : 질문과 문장을 생성 모델에 학습시켜서



딥러닝의 모델의 한계점

① 답변 길이가 긴 경우 성능이 매우 매우 저하



② 모델의 Output들이 모두 Token 단위임

① Seq2Seq

② Transformer

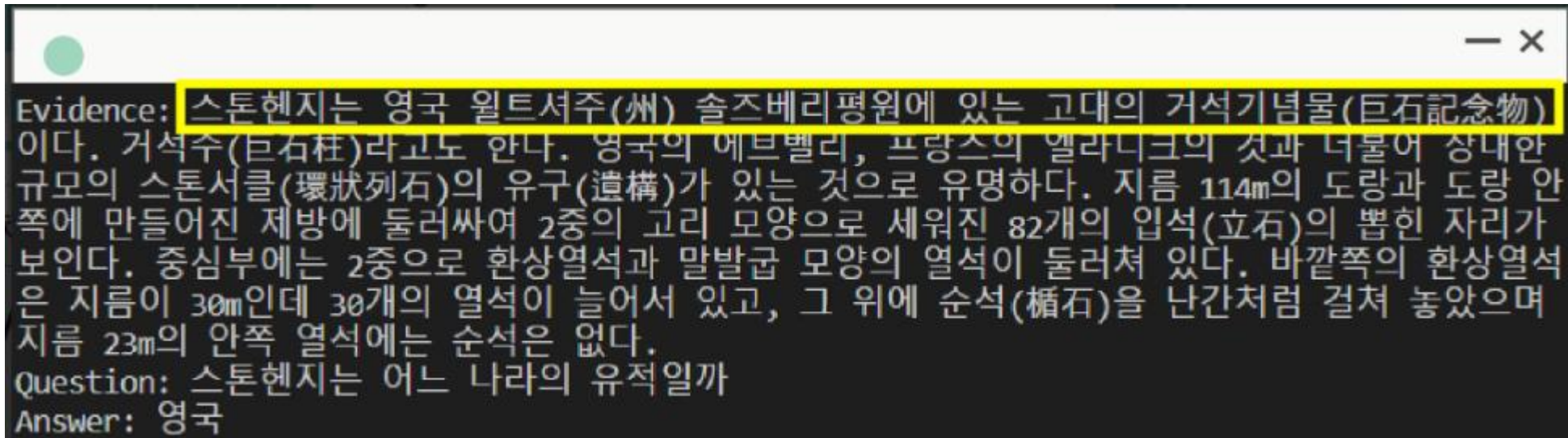
③ BERT

Model

Approach

- 3. Q-A Model : 질문에 대한 답변을 Q-A로 학습시켜 챗봇 답변으로 제시

7회 컨퍼런스 TMI 시연 中



“모델이 전체 문장 중 정답이 존재하여 찾아 가는 방식으로 학습 되므로 우리 데이터 셋과는 맞지 않다고 판단”

Model

Approach

- 3. Q-A Model : 질문에 대한 답변을 Q-A로 학습시켜 챗봇 답변으로

따라서 여러 방법들을 적절히 혼합하여 사용하는 4번째
Ensemble 방법을 고안함.

보인다. 중심부에는 2중으로 환상열석과 말발굽 모양의 열석이 둘러쳐 있다. 바깥쪽의 환상열석은 지름이 30m인데 30개의 열석이 늘어서 있고, 그 위에 순석(楯石)을 난간처럼 걸쳐 놓았으며 지름 23m의 안쪽 열석에는 순석은 없다.
Question: 스톤헨지는 어느 나라의 유적일까
Answer: 영국

“모델이 전체 문장 중 정답이 존재하여 찾아 가는 방식으로 학습 되므로 우리 데이터 셋과는 맞지 않다고 판단”

Model Flow

Step 1:

Input

2급 장애인도
장애인 연금
탈 수 있나요?



사용자의 질문

Model Flow

Step1:

Input

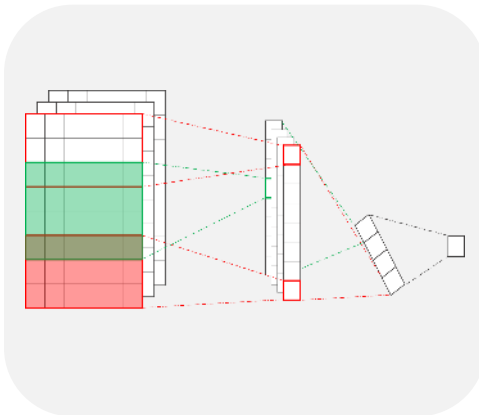
2급 장애인도
장애인 연금
탈 수 있나요?



사용자의 질문

Step2:

Prediction Model



Text
Convolutional
Neural
Network

Model Flow

Step1:

Input

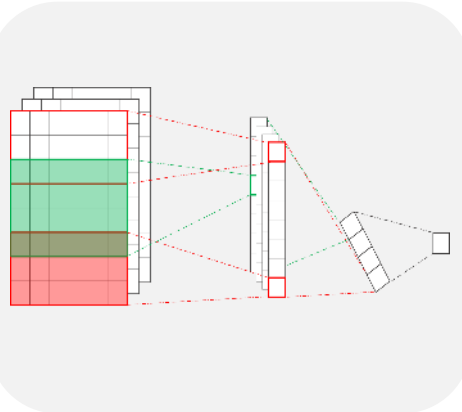
2급 장애인도
장애인 연금
탈 수 있나요?



사용자의 질문

Step2:

Prediction Model

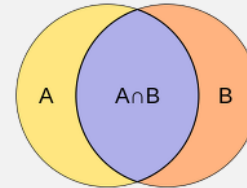


Text
Convolutional
Neural
Network

Step3:

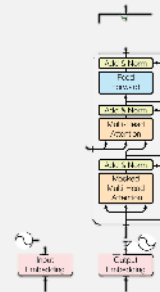
Answer Generation

Class 1 (len > 30)



Jaccard & Cosine
Similarity

Class 0 (len ≤ 30)



Transformer

Model Flow

Step1:

Input

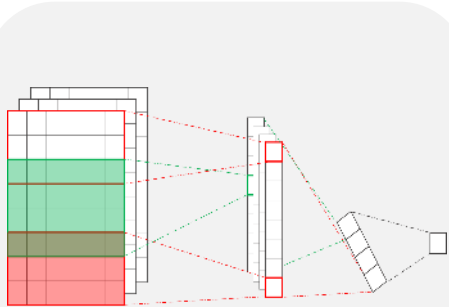
2급 장애인도
장애인 연금
탈 수 있나요?



사용자의 질문

Step2:

Prediction Model

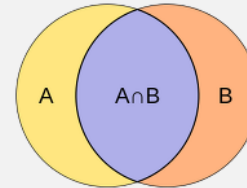


Text
Convolutional
Neural
Network

Step3:

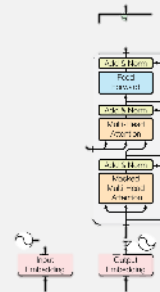
Answer Generation

Class 1 (len > 30)



Jaccard & Cosine
Similarity

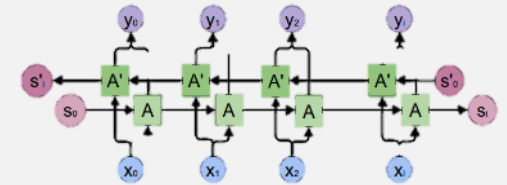
Class 0 (len ≤ 30)



Transformer

Step4:

Post Processing



Bi-Stacked LSTM

Model Flow

Step1:

Input

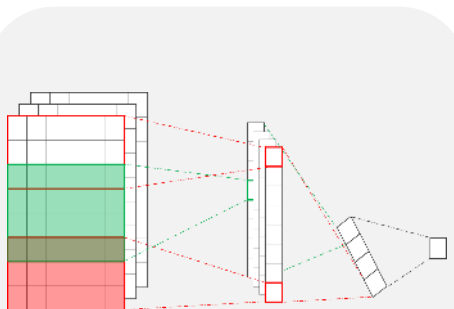
2급 장애인도
장애인 연금
탈 수 있나요?



사용자의 질문

Step2:

Prediction Model

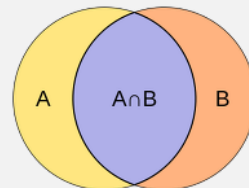


Text
Convolutional
Neural
Network

Step3:

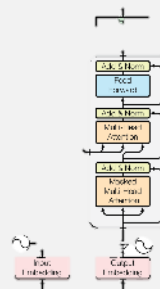
Answer Generation

Class 1 (len > 30)



Jaccard & Cosine
Similarity

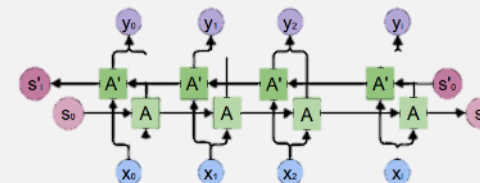
Class 0 (len ≤ 30)



Transformer

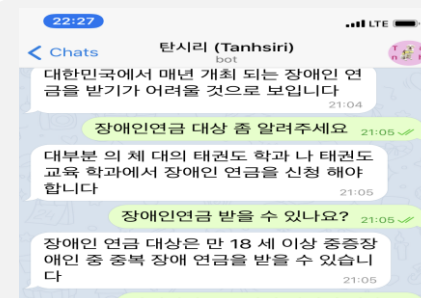
Step4:

Post Processing

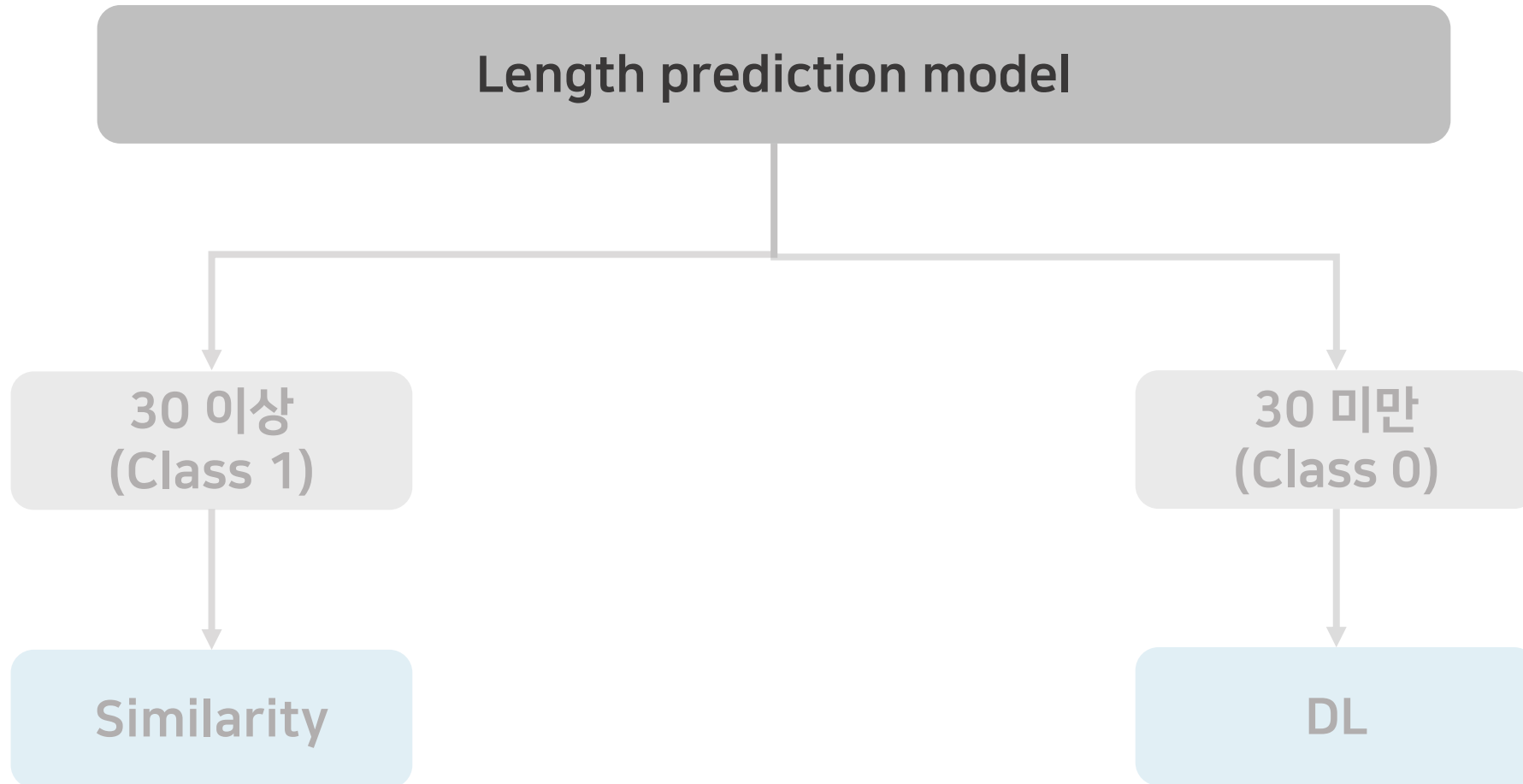


Bi-Stacked LSTM

Step5: Output



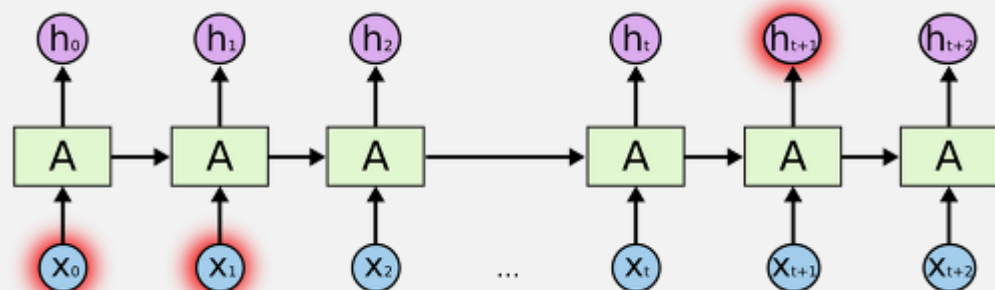
Similarity



Prediction Model (CNN)

(1) Why We Predict a Length of Answer

1) 답변 자체 생성을 시도



→ Long-Term Dependency Problem!

2) 길이 “30” → DL 모델 학습 적절한 수준

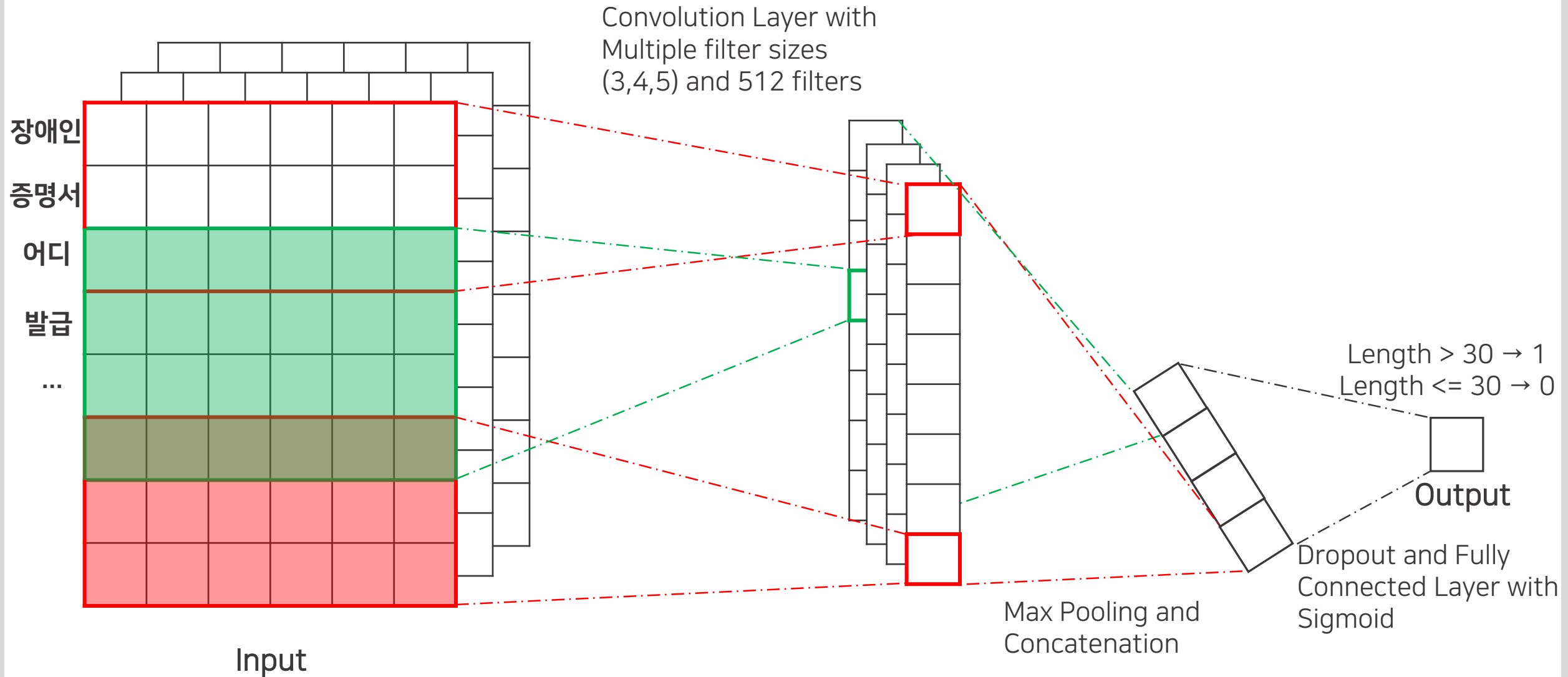
(2) Why CNN for Length Prediction?

(Accuracy)

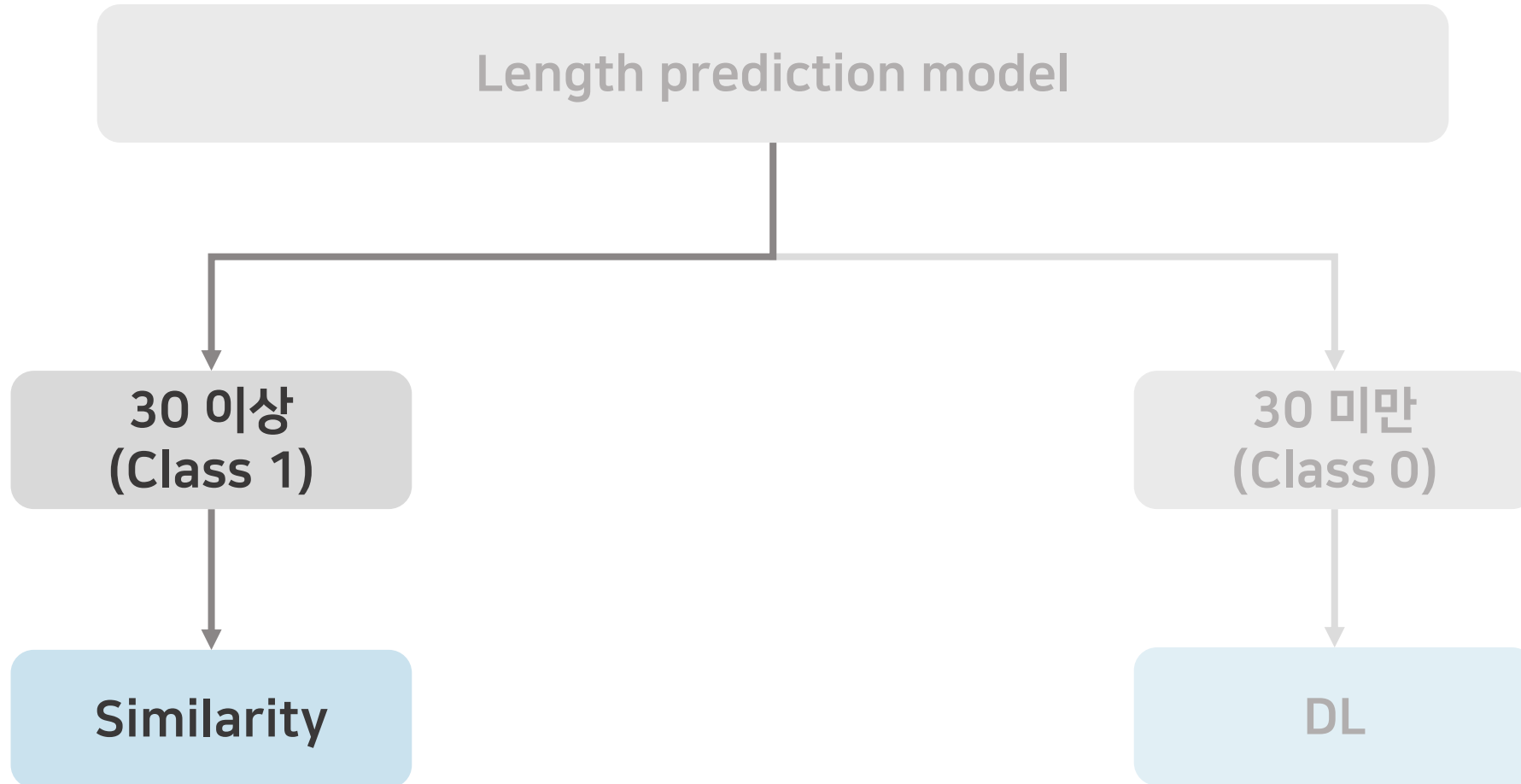
Machine Learning			Deep Learning	
Random forest	XGBoost	lightGBM	CNN	RNN
0.8029	0.8134	0.7961	0.9601	0.9593

DL이 ML보다 뛰어난 성능을 보이고 있음

Prediction Model (CNN)



Similarity



Similarity

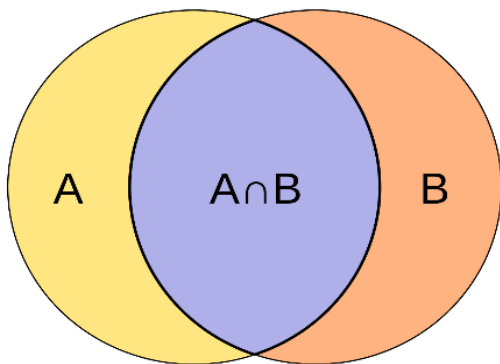
STEP1. 새로운 질문 Input으로 부터 Dataset에 있는 질문과의 유사도를 구함.

STEP2. 유사도의 경우 아래와 같은 방식으로 계산을 하였음.

STEP3. 가장 유사도가 높은 질문의 답변을 Chatbot의 답변으로 제시

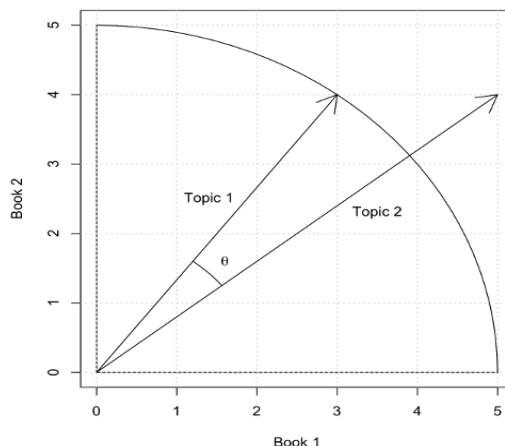
Jaccard

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$



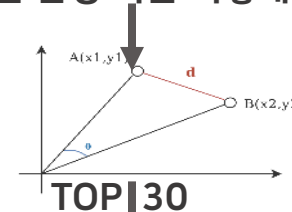
Cos similarity

Doc2Vec embedding



Ensemble

Q. 장애인 판정 기준 어떻게 되나요?



Sim_Q. 장애인 판정 받으려면 어떤 기준이 있나요?

A. 장애인 판정 기준은 다음과 같습니다. (...)

Advanced Similarity

유사도를 계산할 때 각 Question의 topic을 파악하여 활용하자

Trial1 LDA Topic modeling으로 주제를 뽑고자 함 → 😞

Trial2 '장애인' token 주변 단어가 중심 주제일 것이다

Ex1) 혹시 고속버스도 장애인 할인 되나요?

Ex2) 장애인 연금 어떻게 받나요?

Advanced Similarity

장애인 특별

장애인 택시

장애인 차량

장애인 소득공제

장애인 주차

장애인 성폭행

장애인 증명서

장애인 인권

장애인 등급

장애인 일

장애인 복지

장애인 수급

장애인 석

장애인 카드

장애인 공무원

장애인 판정

...

등의 82개의 topic 선정하여
해당 topic에 해당하는 글로만
유사도를 계산

Advanced Similarity

Length prediction model

30 이상

Question에서 topic 찾기

Topic token 존재 O

Topic token 존재 X

동일한 topic의 Q 중 가장 높은 유사도를 구함

Total Q 중 가장 높은 유사도를 구함

threshold

threshold

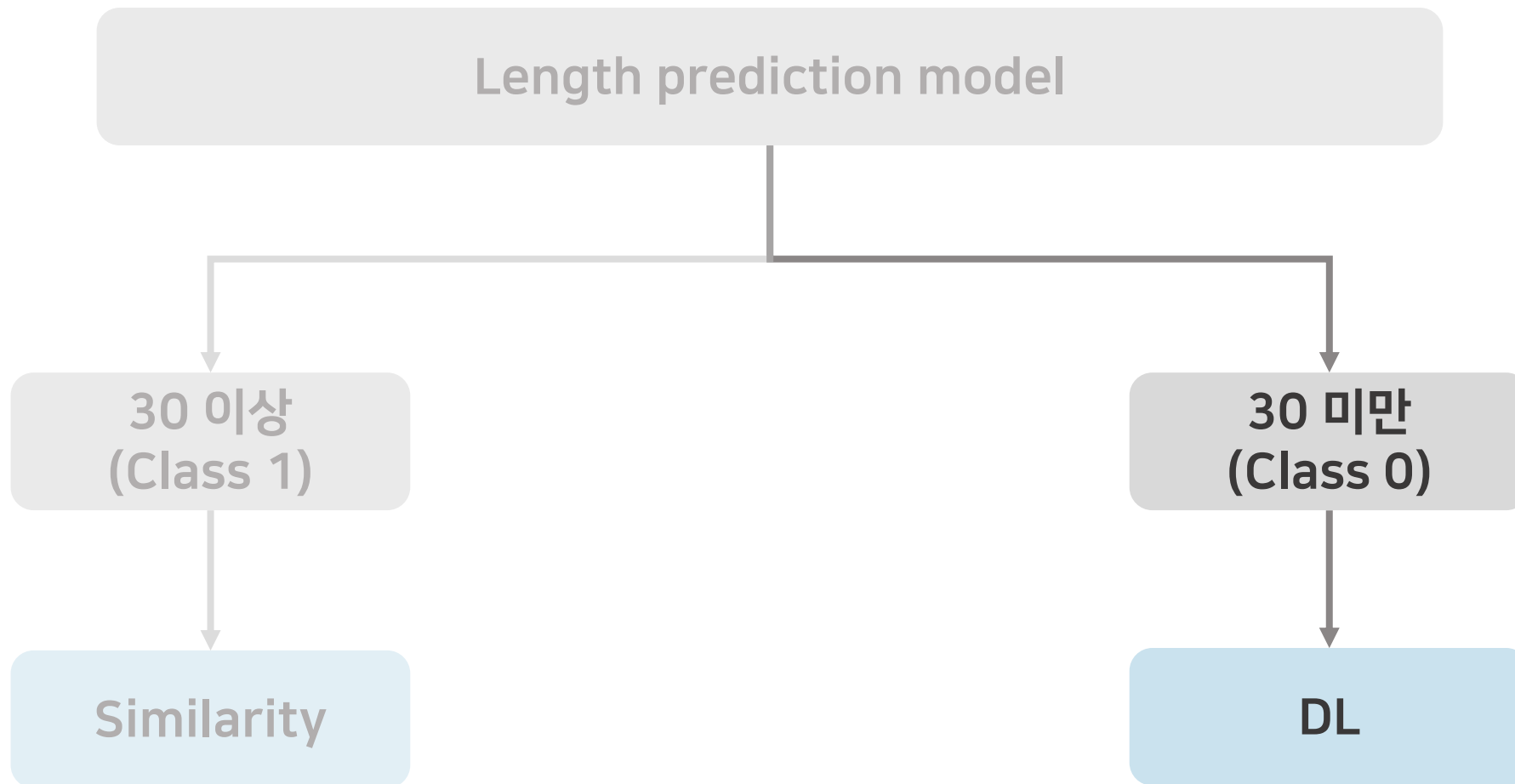
Sim_Q에 대응되는 A

주제에 대응되는 link

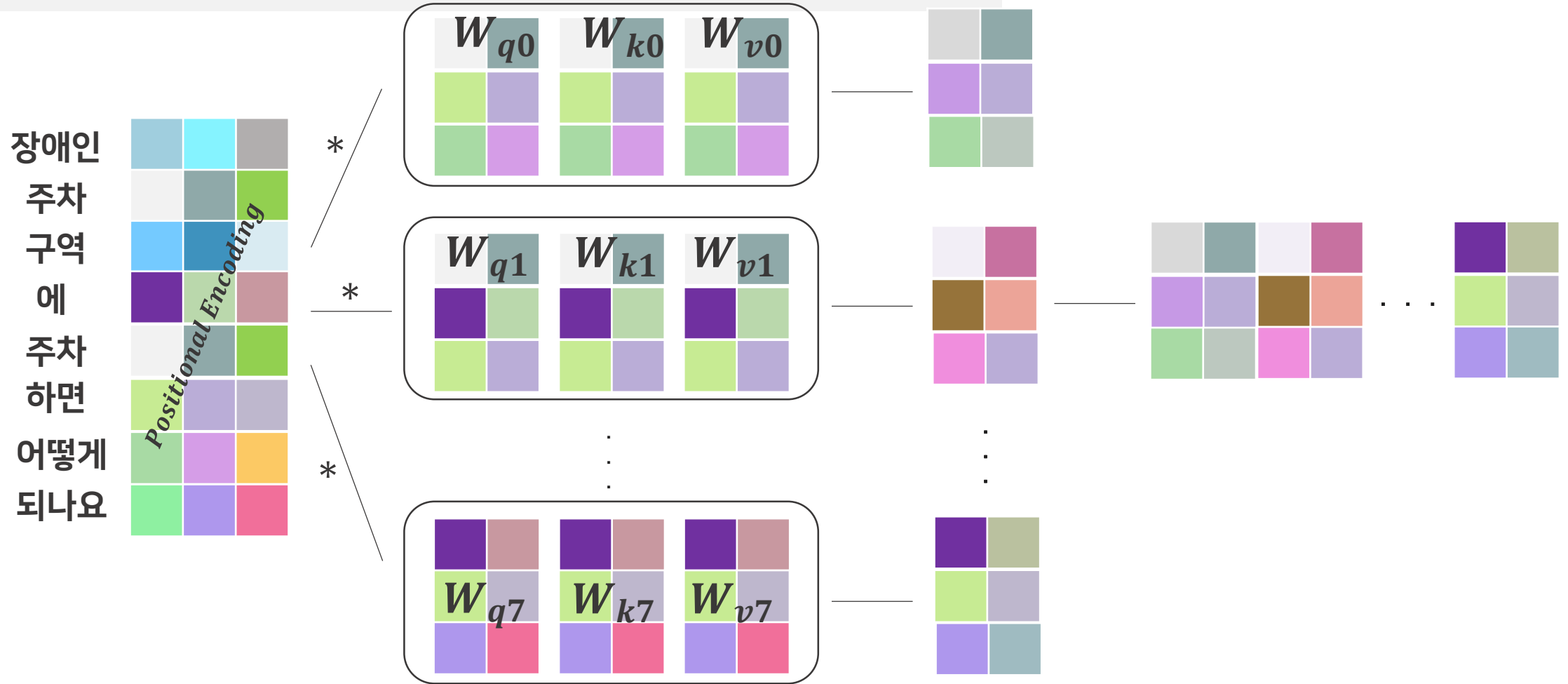
Sim_Q에 대응되는 A

재질문 요청

Transformer

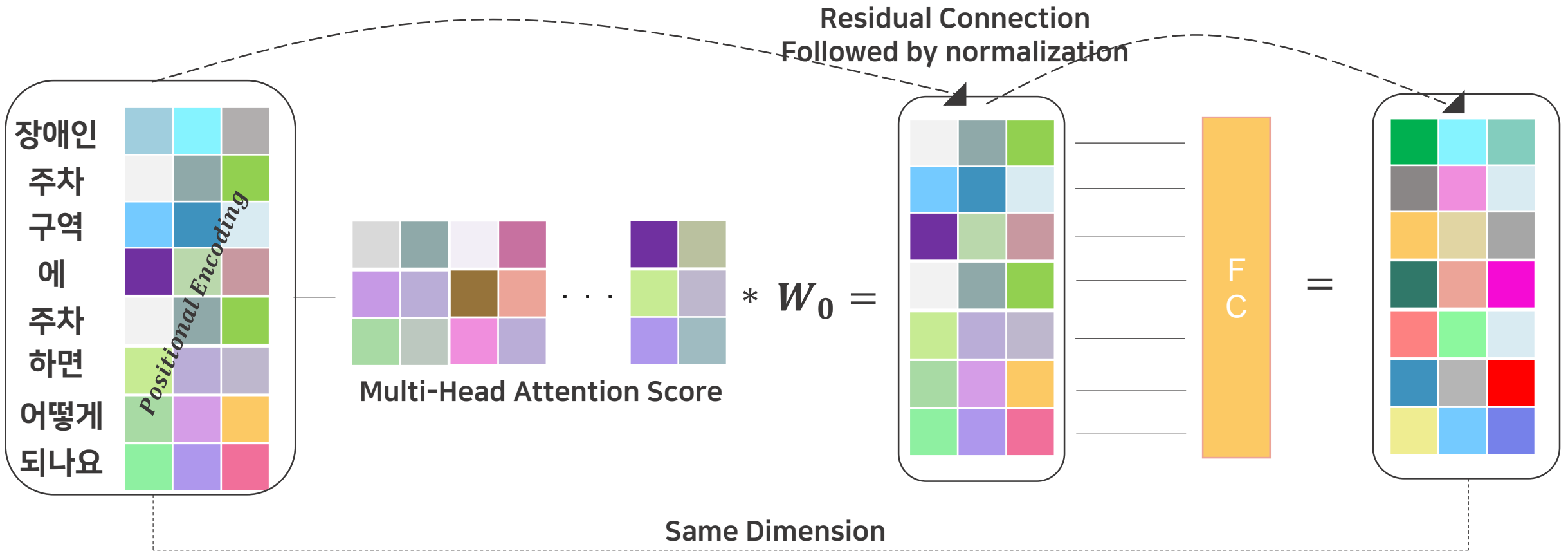


Transformer : Encoder



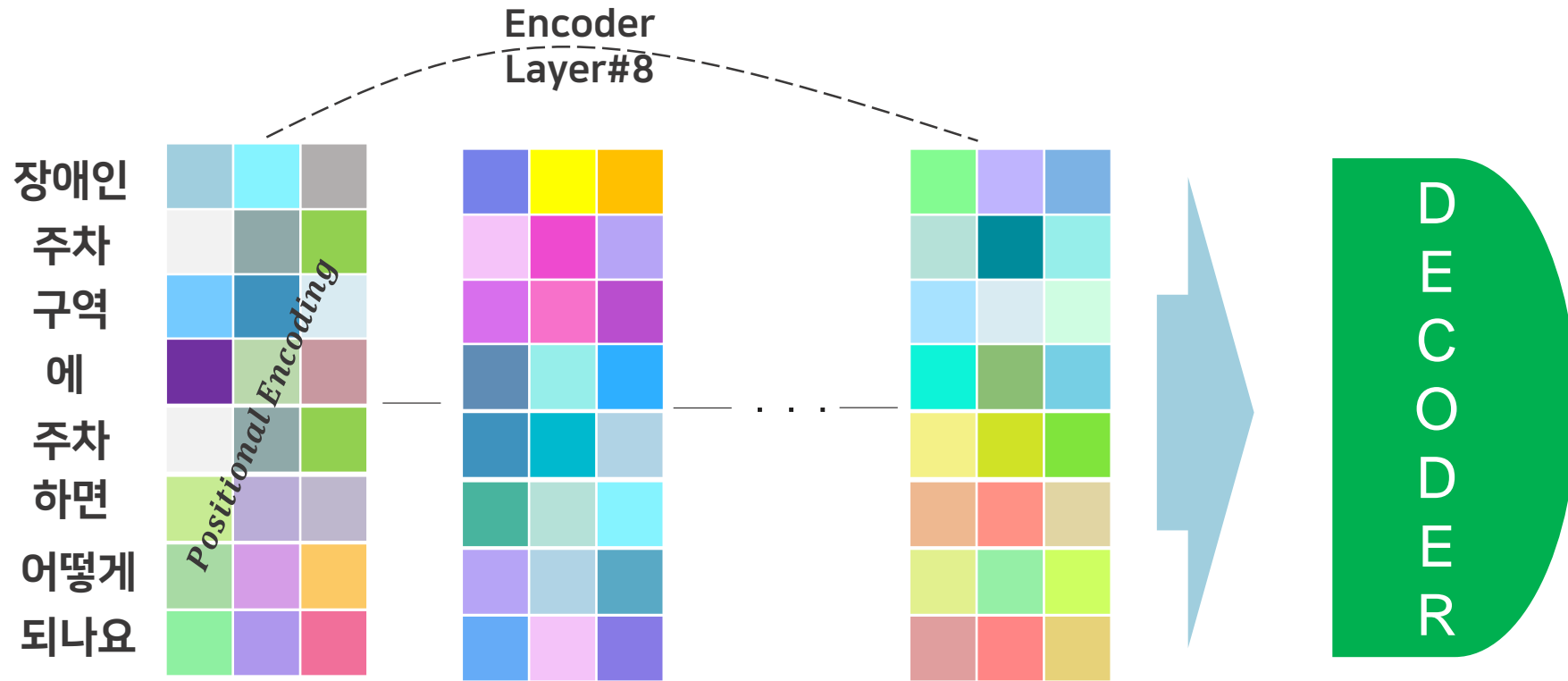
여러 개의 Query, Key, Value의 Weight을 병렬적으로 학습해 Multi-Head Attention이 수행됨.

Transformer



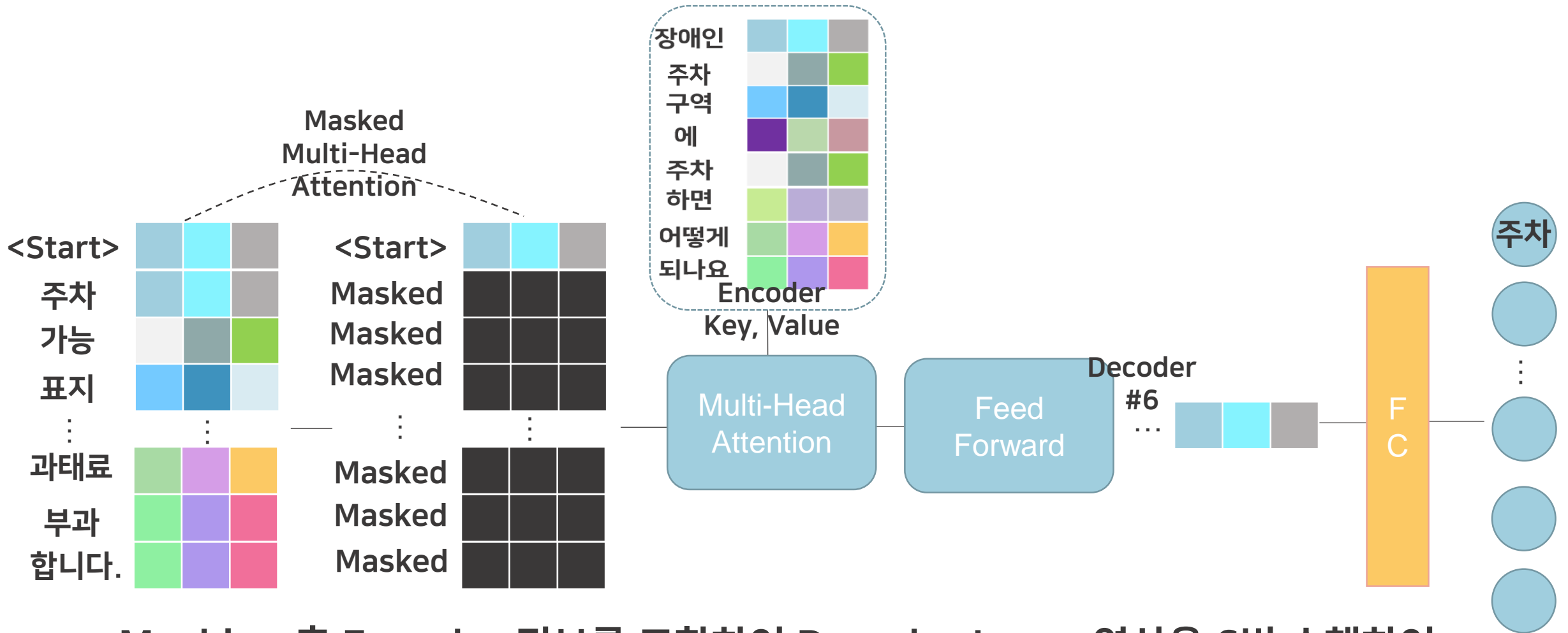
해당 Attention Score의 한번의 행렬 곱과 FC를 통해 처음 Dimension 같은 벡터를 생성함. 또한 처음 정보를 잊지 않도록 Residual Connection을 수행

Transformer : Decoder



이러한 Encoder Layer 연산을 8번 수행하여 Decoder에 전달함.

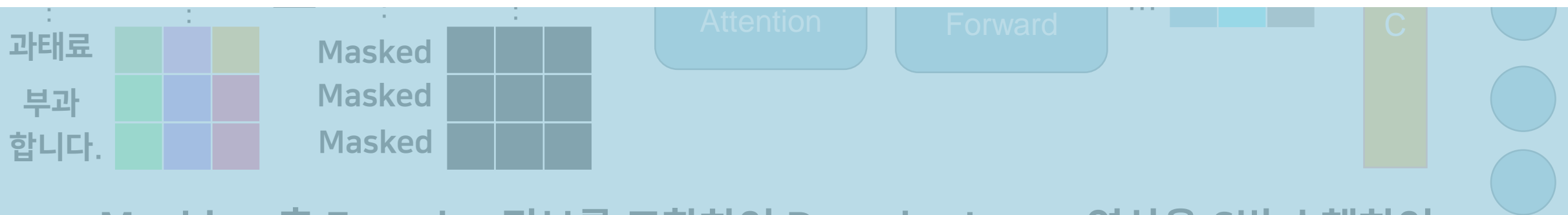
Transformer : Decoder



Masking 후 Encoder 정보를 포함하여 Decoder Layer 연산을 6번 수행하여 이를 최종적으로 FC를 통해 다음 Token을 예측해 나가는 방식으로 학습이 진행됨.

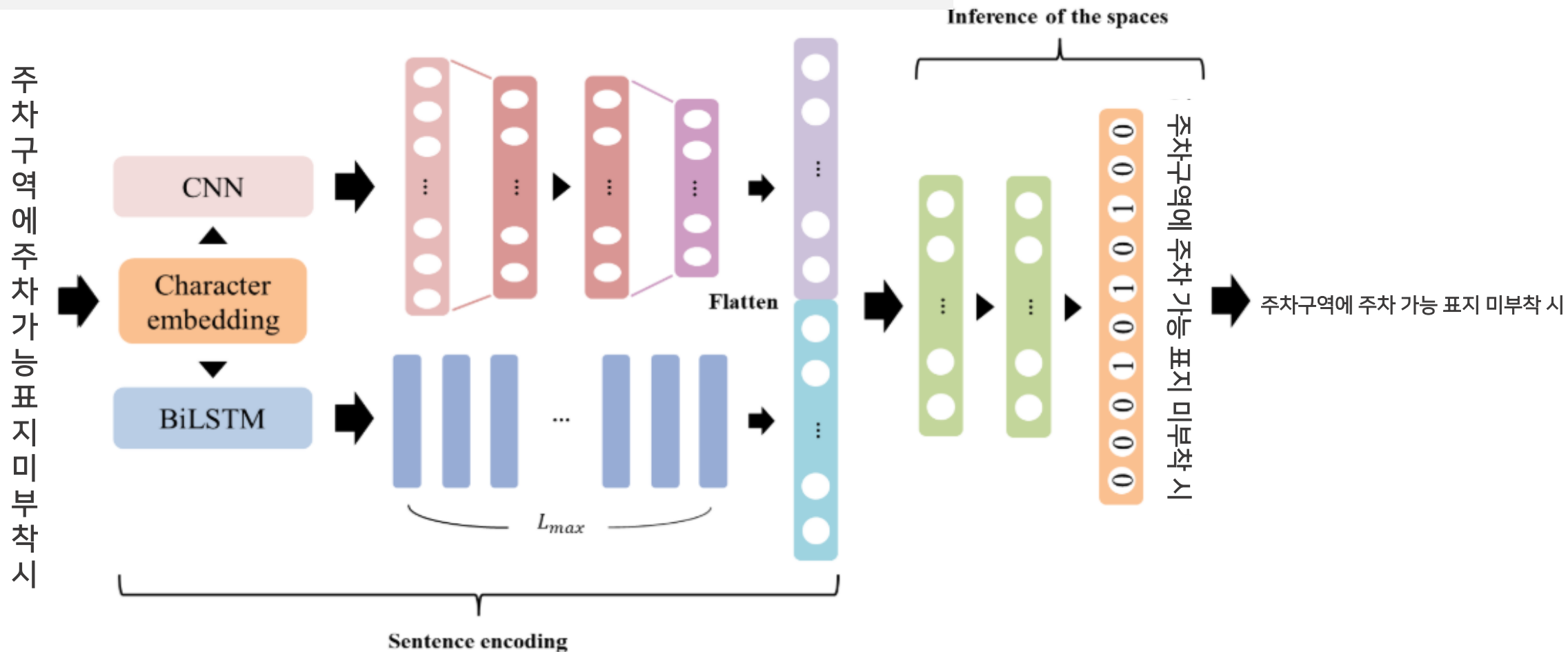
Transformer : Decoder

유사도의 경우 질문자의 말투에 민감하게 반응하여 같은 주제임에도 불구하고 다르게 답변되나, 딥러닝의 경우 이러한 문제에 대해 Robust하게 답변이 가능함.



Masking 후 Encoder 정보를 포함하여 Decoder Layer 연산을 6번 수행하여 이를 최종적으로 FC를 통해 다음 Token을 예측해 나가는 방식으로 학습이 진행됨.

Postprocessing



CNN의 Feature와 BiLSTM의 Feature를 concat한 후 이를 FC를 통해 음절 별 띄어쓰기가 이루어지는지를 예측

Postp

Rule Based

장애인도 신체검사 받나요 오전 10:14 ✓✓

장애인은 병역이 면제 되나 장애인 등록자 중 병역 판정 검사 대상자의 경우 병역 판정 검사를 받아야 합니다
오전 10:14

Deep Learning

장애인도 신체검사 받나요 오전 11:39 ✓✓

장애인은 병역이 면제되나 장애인 등록자 중 병역 판정 검사 대상자의 경우 병역 판정 검사를 받아야 합니다
오전 11:52

CNN의 Feature와 BiLSTM의 Feature를 concat한 후 이를 FC를 통해 음절 별 띄어쓰기가 이루어지는지를 예측

시연

1) 답변 길이가 긴 경우 -> Similarity

Case 1 : 존재하는 키워드 내에서 test

Case 2 : 존재하는 키워드 내에서 test

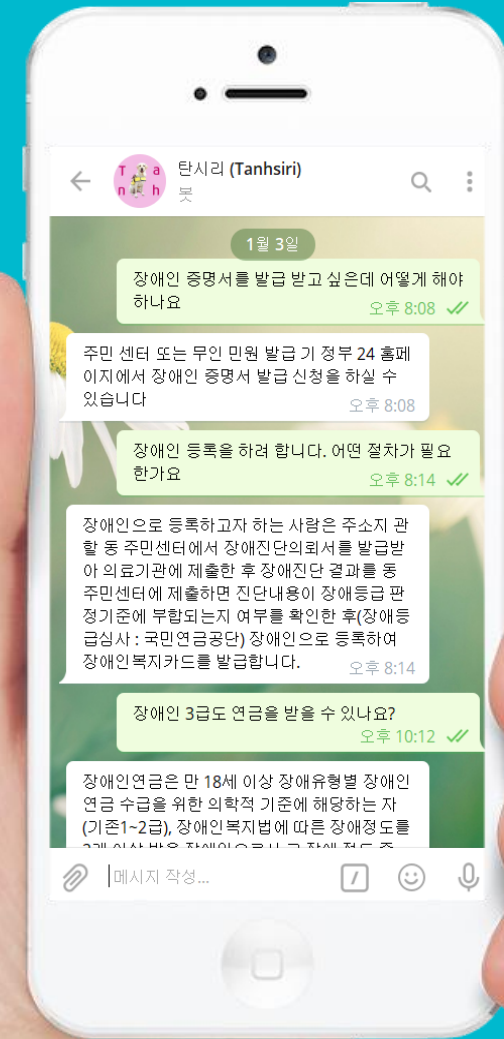
But, 유사도가 낮은 경우

-> 관련 기관의 링크를 제공

Case 3 : 존재하지 않는 키워드 내에서 test

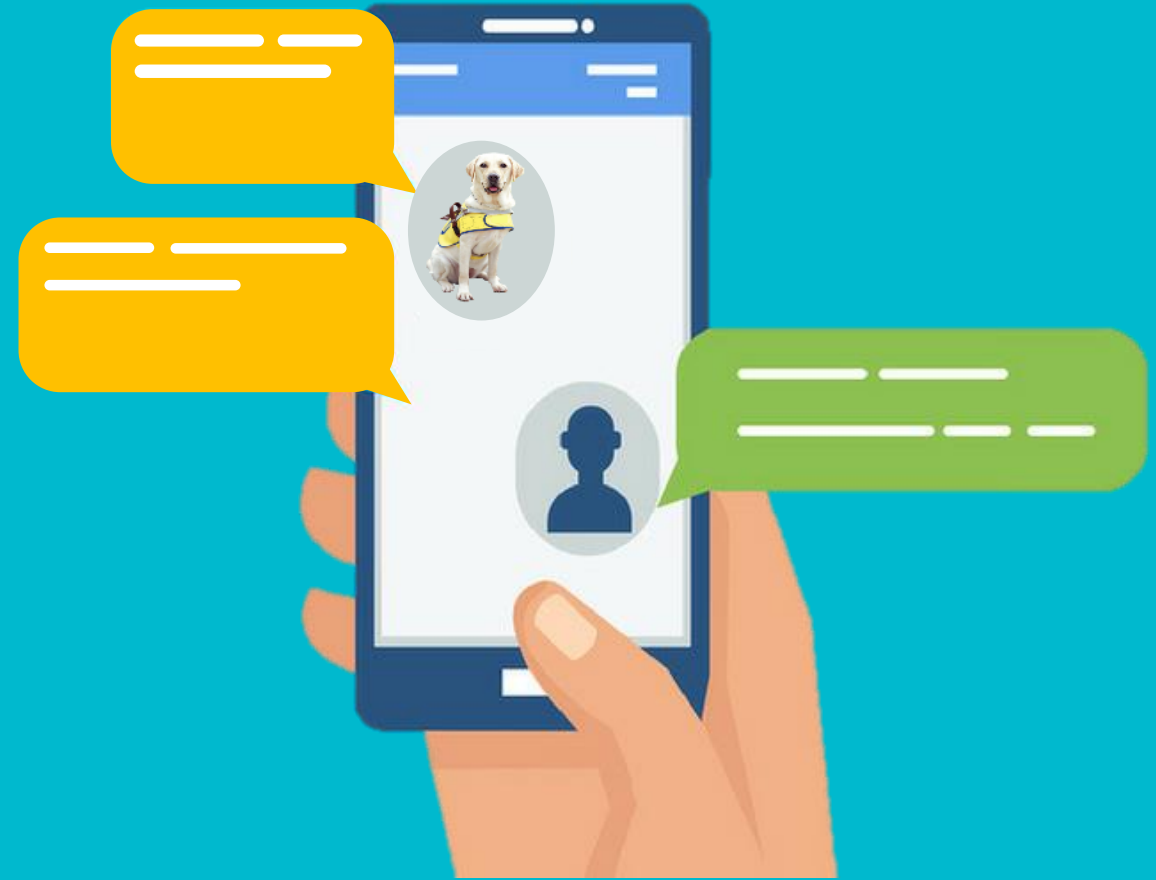
-> 답변 불가

◆ 답변 길이가 짧은 경우 -> 딥러닝



Part5.

기대효과 및 한계점



기대효과

1. 보다 쉽고 빠른 정보 접근

- 포털 사이트 검색 혹은 관련 기관 문의에 소요되는 불필요한 대기 시간 단축
- 별도의 절차 없이 메신저 채팅만으로 손쉽게 정보를 얻을 수 있음

장애인연금 인상 안됐나요.?	장애인복지	0	12시간 전
50 장애인활동지사	장애인복지	1	12시간 전
100 자동차 이전 및 장애인주차표찰 발부	자동차세	0	13시간 전
100 시각장애인 외국 방문시 출입국심사 관련...	외국법	0	22시간 전
50 장애등급 폐지되면 경증 정신장애인으로...	장애인복지	1	22시간 전

장애인연금 조건이 어떻게 되나요?

21:06 ✓

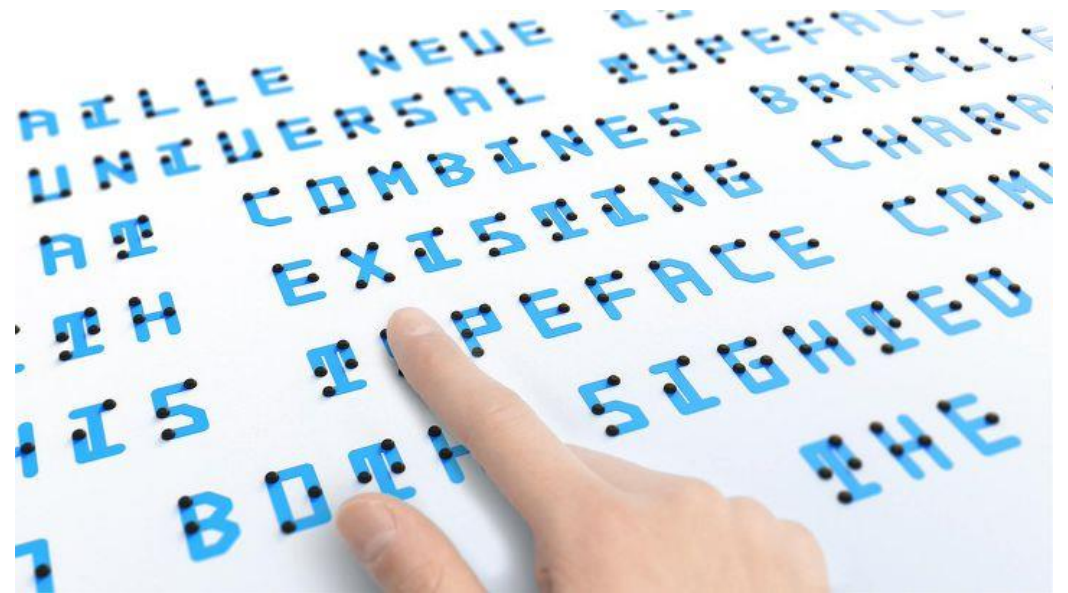
장애인 연금은 만 18 세 이상 중증장 애인
중 소득 인정 액 이 선정 기준 액 이하 인
자에게 지급 됩니다

21:06

기대효과

2. 다양한 platform과의 연동

- 현재는 text이지만 이를 엔진으로 활용하여 음성, 점자 등으로 확장 가능



기대효과

3. 관련 기관과의 연계

- 장애인 복지 관련 기관과의 연계를 통해 보다 더 신뢰성 있는 정보 제공



장애인활동지원



한국장애인복지시설협회
Korea Association of Welfare Institutions for Persons with Disabilities

한계점



데이터 셋에 존재하는
질문만 답변 가능

지식인, 다음 팁 데이터로 답변
데이터 구축



지엽적인 사연에 대한
답변의 어려움

질문이 매우 지엽적일 경우
미흡한 답변 발생 가능



질문자의 어투에 민감
in similarity

유사도 기반의 답변의 경우
질문자의 어투에 따라 답변이
민감하게 변화

QnA

