



# 딥러닝 법률상담 AI

Best-m4.t3

---



© 2023 All Rights Reserved. All Rights Reserved.

# 목차

01

개요

프로젝트 배경

02

데이터 전처리

네이버 지식인 크롤링  
데이터 전처리  
Textrank

03

모델링

파이프라인  
군집화  
유사도  
워드 임베딩

# 목차

04

모델링

Seq2Seq  
Attention  
KoGPT2

05

기대효과 및 한계점

접근성 / 가능성  
데이터 / 자원

06

시연 및 Q&A

Flask 시연  
Q&A



# 팀원소개

## 신재웅

- 프로젝트 총괄
- 데이터 크롤링
- 모델검증

## 박찬규

- 최종발표
- 데이터 크롤링
- 알고리즘 아키텍처 담당

## 권기일

- 깃허브 관리
- 데이터 크롤링
- 모델구성

## 강신웅

- 기획안 발표
- 데이터 크롤링
- 데이터 전처리

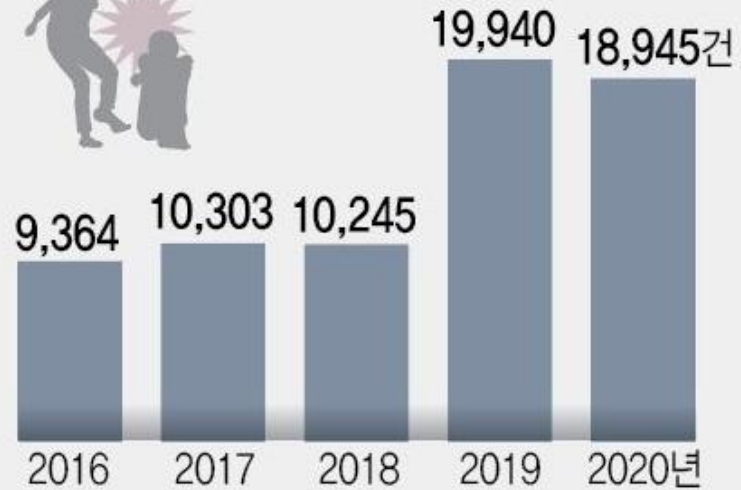
# 01

## 프로젝트 배경

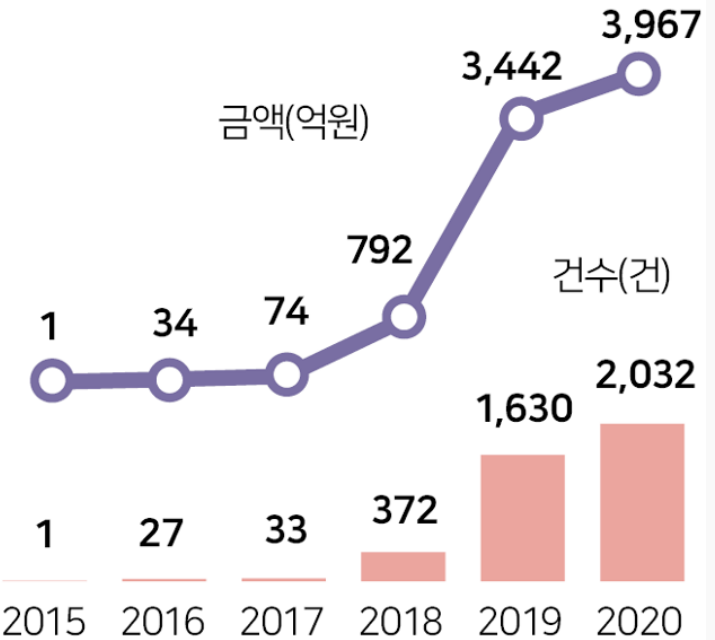
일상생활의 법률 사건사고 증가

# 01. 일상생활 법률 사고 증가

## 데이트폭력 신고건수 현황



## 전세보증금반환보증 보증사고



●2020년은 1~10월 합계, 자료 HUG

# 02

## 데이터 전처리

네이버크롤링 및 텍스트랭크

# 01. 네이버크롤링



NAVER 지식iN

Q

전문가라면 엑스퍼트로 모

홈

Q&A

답변하기

지식기부

사람들

베스트

명예의전당

프로필

Q&A

> 사회, 정치

> 법, 법률

+

법, 법률

오늘의 새 질문 1,907

모바일 질문 1,277

오늘의

교통 사고, 위반 548,349

부동산, 건축 584,969

신용, 파산 367,576

가족, 이혼 429,366

지식재산권 278,268

형사사건 2,042,806

소비자관련법, 상법 258,104

청소년관련법 180,157

@ 세금, 세무

재판, 소송 절차 545,170

계약 260,747

손해배상 328,480

형벌, 형집행 460,228

법학, 법이론 283,051

언론, 미디어 관... 128,085

외국법 102,202

헌법 31,292

행정법 83,047

민법 135,769

민사소송 181,668

노동법 144,664

등기 61,692

민사집행 86,906

산업재해 198

- 훈련용 데이터셋 : 15,000건
- 검증용 데이터셋 : 3,000건

NAVER 지식iN

Q

전문가라면 엑스퍼트로 모십니... exp

홈

Q&A

답변하기

지식기부

사람들

베스트

명예의전당

프로필

파트너센터

Q&A

> 사회, 정치

> 법, 법률

> 민사소송

+

민사소송

오늘의 새 질문 18

모바일 질문 13

오늘의 답변 39

질문

법률 전문가 답변 질문

랭킹

제목

분야

작성자

작성일

명도소송 전에 보내는 내용 증명에 넣어야 하는 내용 질문.

민사소송

정종욱

19분 전

민사재판 항소

민사소송

이종결

44분 전

변호사선임료

민사소송

서동근

11시간 전

**엄정숙 변호사님 답변**  
**법도 종합 법률사무소** · 받은감사수 31

안녕하세요. 서울지방변호사회-네이버 지식iN 상담변호사 엄정숙 입니다. 대한변호사협회에 '부동산', '민사' 전문분야로 공식등록(제2013-72호, 제2013-71호)되어있는 '부동산 전문변호사', '민사 전문변호사'입니다.

명도소송 중에 연체차임과 관리비를 모두 지급하여 다시 임대차계약을 유지하는 경우에는  
장래 계약해지사유 발생시 명도한다는 내용으로 현재 소송에서 조정이나 화해권고 등으로 판결을 받아두는 방법이 유리하실 것으로 보입니다.

08



## 02. 데이터 전처리

+



김부조 변호사님 답변

법무법인 해울 · 받은감사수 3

안녕하세요. 서울지방변호사회-네이버 지식iN 상담 변호사 김부조입니다.

사해행위 취소소송의 경우 사해행위를 안날로부터 1년, 있는 날로부터 5년의 제척기간이 있는바, 귀하의 질문대로 2016. 8.월 채무자 명의의 아파트가 배우자 명의로 변경이 되었다고 하더라도 이미 제척기간이 도과하였는바, 사해행위 취소소송을 제기하더라도 승소하지 못할 가능성이 높다 할 것입니다.

보다 자세한 상담을 원하시는 경우 네임카드 기재 연락처로 연락주시면 상담도와드리겠습니다.

감사합니다.



김부조 변호사

법무법인 해울 · 02-6925-4576

지식iN 변호사 답변은 대국민 법률 상담에 뜻을 둔 변호사의 지식기부로서 서울지방 변호사회와 네이버의 제휴를 통해 변호사가 직접 정확하고 신뢰할 수 있는 답변을 ...

예약

홈페이지

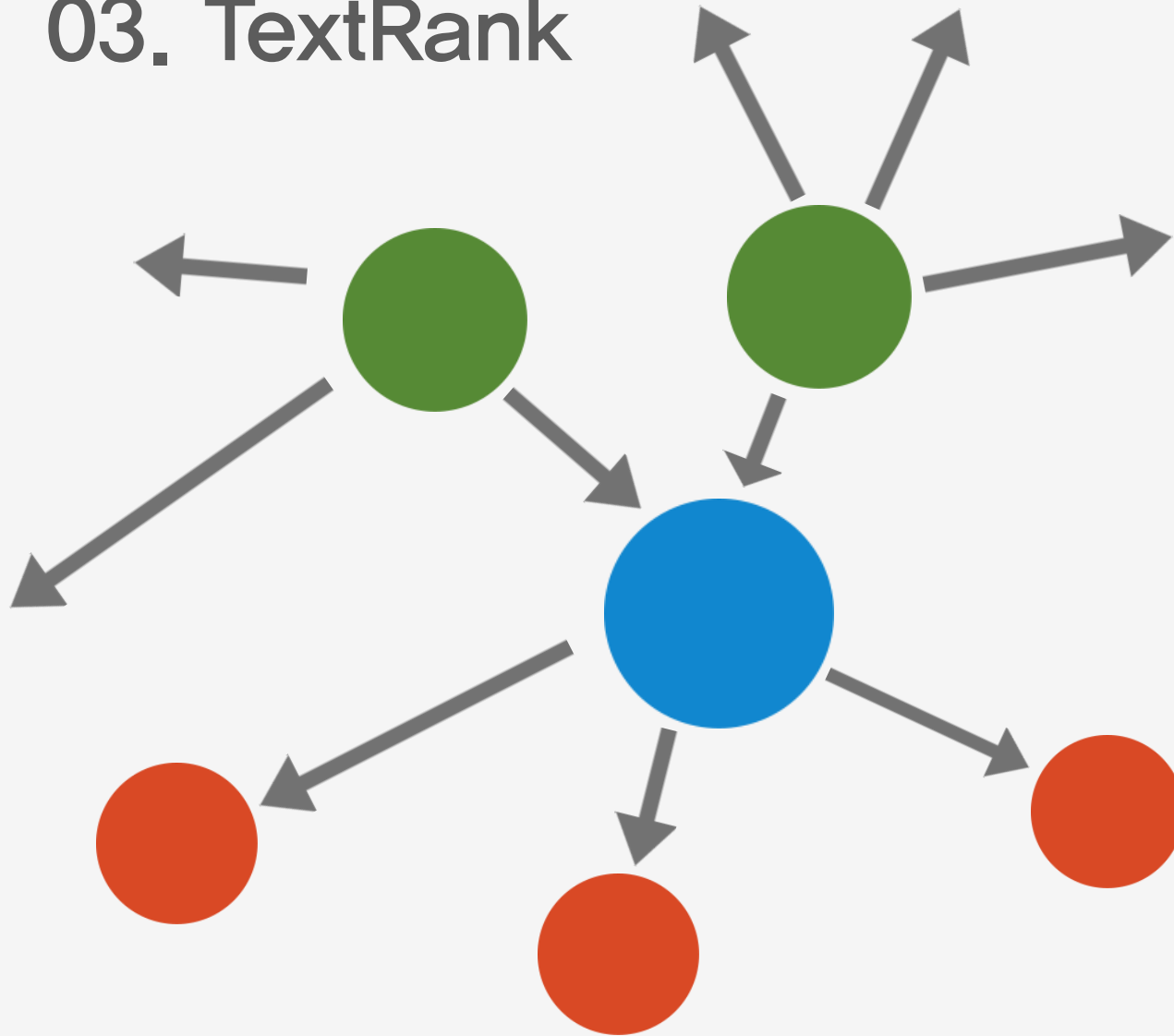
위치

( 불필요한 인사말 )

필요한 답변  
추출!

( 홍보문구 )

## 03. TextRank



- 1) 각 노드는 균등한 가중치에서 시작
- 2) 연결된 노드로부터 가중치 전달
- 3) 재귀적 학습으로 가중치 업데이트
- 4) 최종 가중치 결과를 랭킹
- 5) 랭킹에 따라 중요한 단어 선택

= 연결된 단어의 수가 많을수록,  
문서 내에서 해당 단어가 자주  
등장할수록 **랭킹 Up!**

-> 최대 5문장 추출!

## 03. TextRank

1원고는 대출을 받아 모든 금액을 피고에게 빌려주었고  
대화 내용으로 이자 처리 가능하겠냐는 말에 수금 후  
총 다섯 차례 변제하였지만 그 이후 연락 두절되어 변제하지 않고 있는 상황입니다.

**이 상황에서 대출이자로 인한 피해 금액과 대여금을 포함한 금액을  
어떻게 산정하여 제출해야 하는지 도움 주시면 감사하겠습니다.**

2 공시송달을 보내었는데 이후 변론 기일에 판결이 나면  
추가로 작성해야 하는 서류가 있다면 무엇이 있을까요.



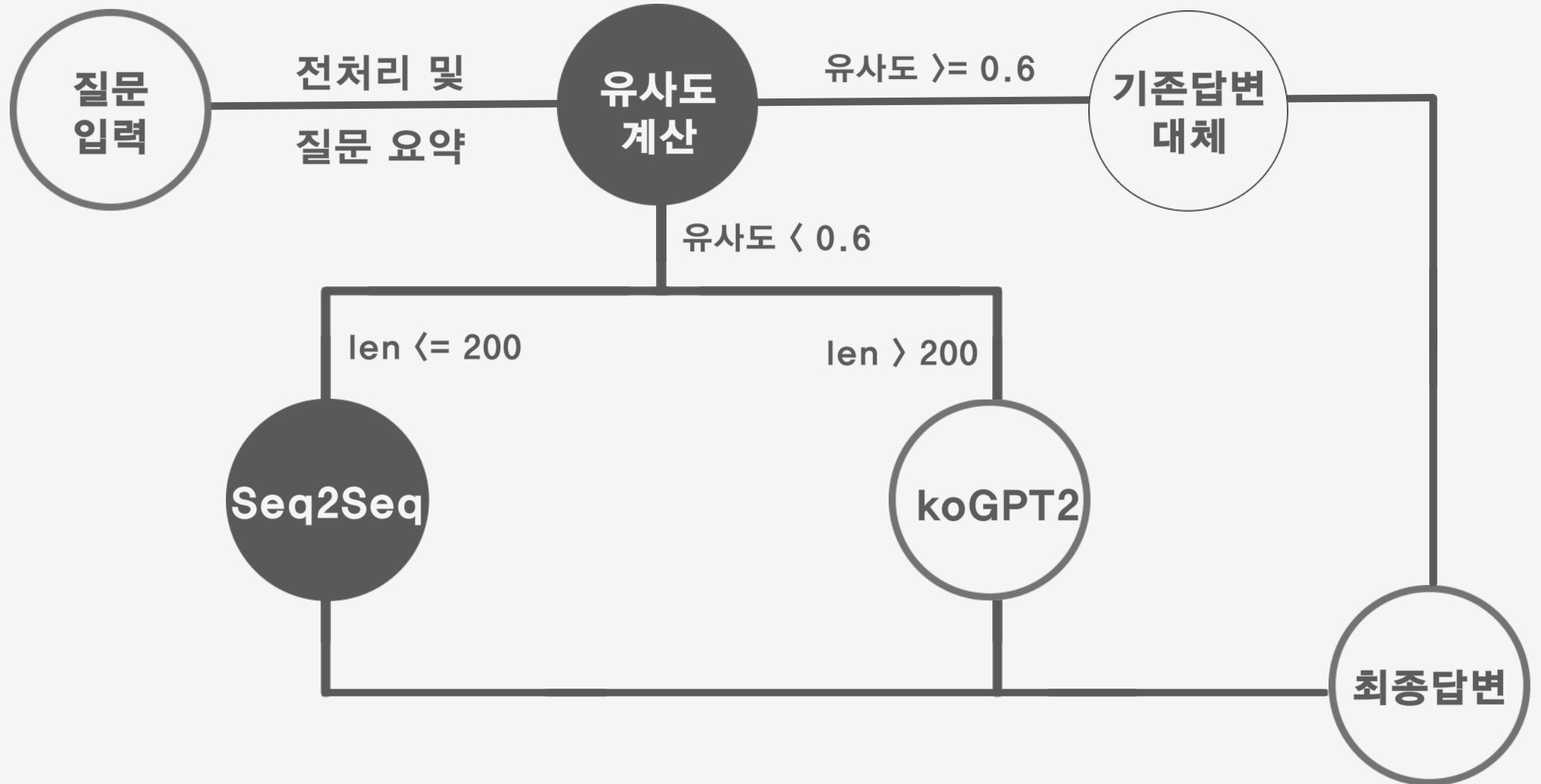
**이 상황에서 대출이자로 인한 피해 금액과 대여금을 포함한 금액을  
어떻게 산정하여 제출해야 하는지 도움 주시면 감사하겠습니다**

# 03

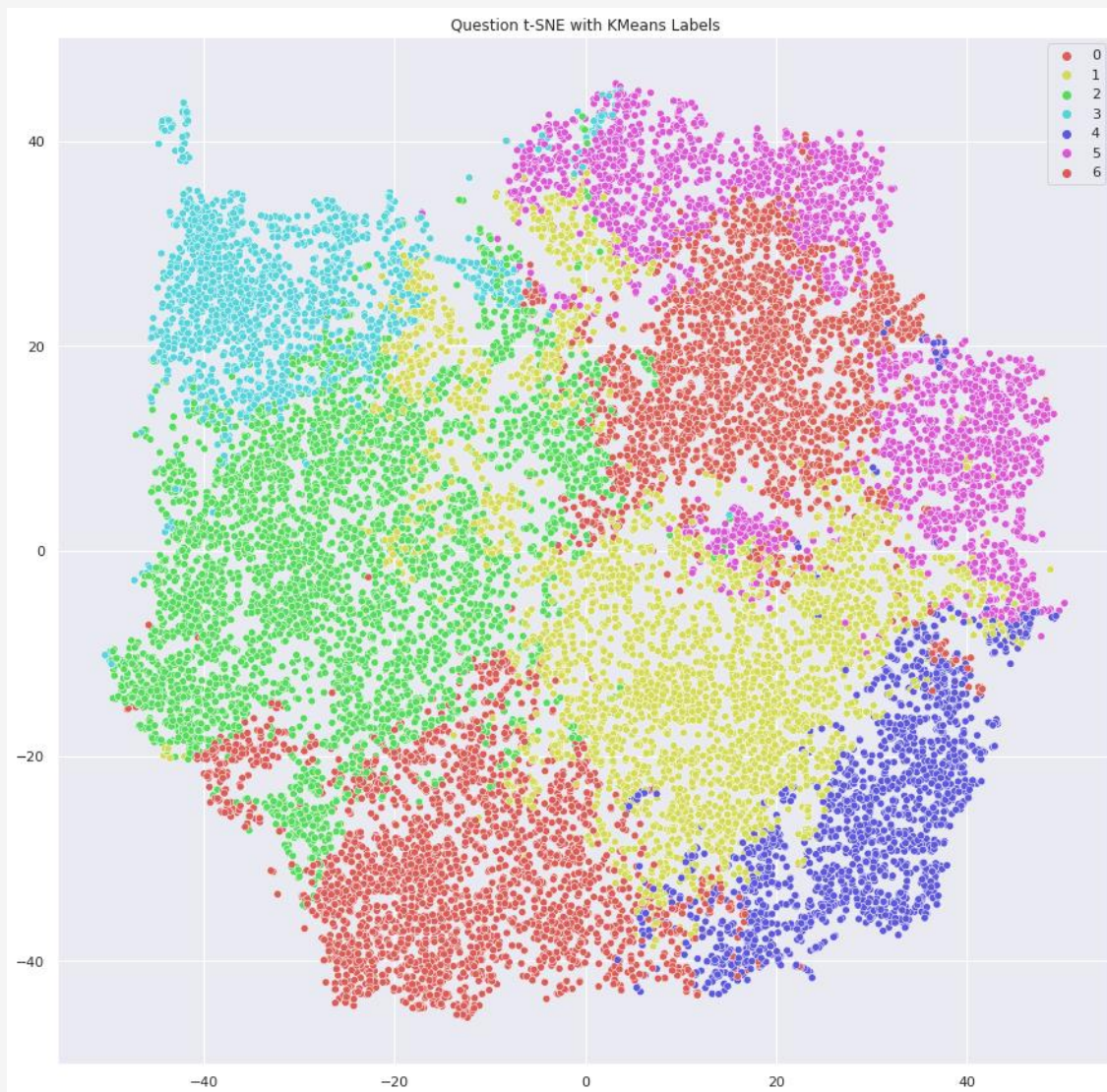
## 모델링

파이프라인 / 군집화 / 유사도 / 워드 임베딩

# 01. 파이프라인



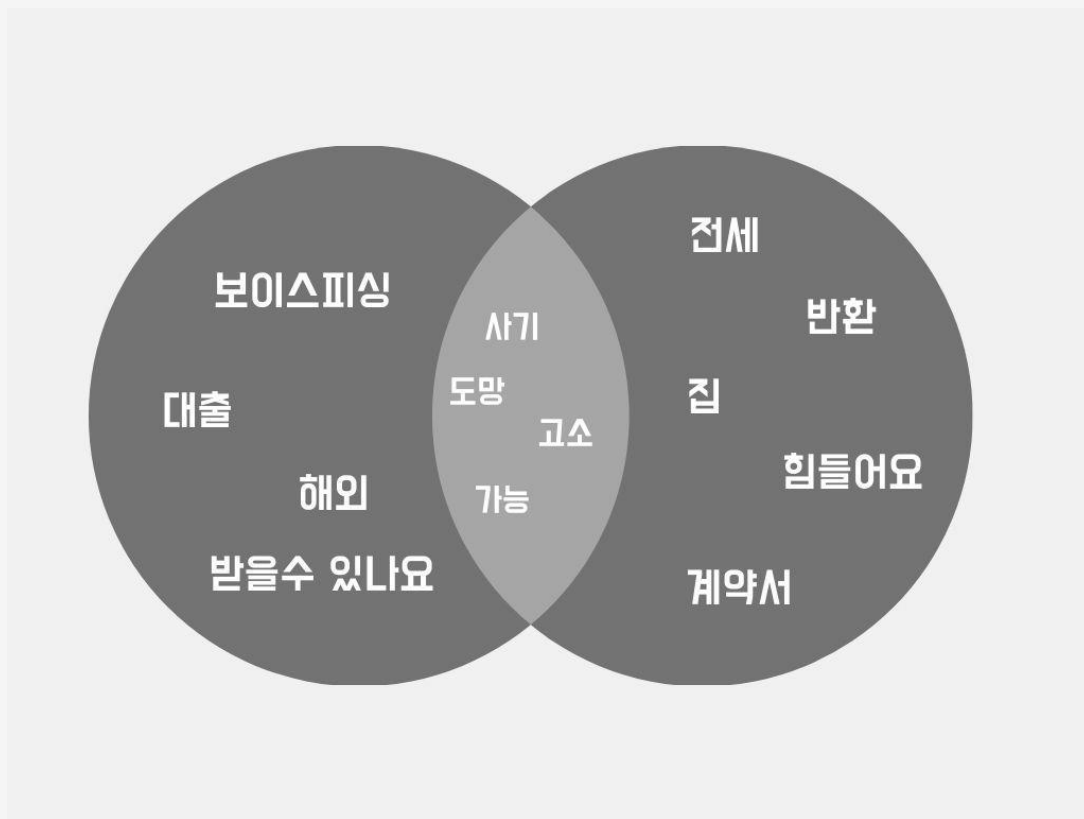
## 02. 군집화



– 단어 단위 임베딩 모델 GloVe를  
이용하여  
문장을 벡터화하고  
클러스터링

– 클러스터링 결과,  
비슷한 질문들의 클러스터 발견  
→ 유사도 모델을 사용!

## 03. 자카드 유사도



자카드 유사도 :

두 문장을 단어의 집합으로 만든 뒤,  
전체 단어를 바탕으로 교집합의 크기를  
합집합의 크기로 나눔

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

교집합: 4

합집합: 13

=> 자카드유사도:  $\frac{4}{13}$  ( $\approx 0.3077$ )

두 문장 사이의 단어가 많이 겹칠수록 커짐

## 03. 자카드 유사도

자카드 유사도는 2가지 문제점 :

### 1) Term Frequency :

집합 단위 계산으로 인해 빈도를 고려하지 못함

### 2) Document Frequency :

해당 단어가 문서에 어떻게 분포되어 있는지 고려하지 못함

Ex) 모든 문서 내에서 발견되는 단어는 유사도가 높지만, 흔하기 때문에 중요하지 않음

→ 같은 단어라도 중요도가 달라질 수 있는데 자카드 유사도는 이를 간과

**TF-IDF : Term Frequency와 Document Frequency를 이용하여 가중치 부여**

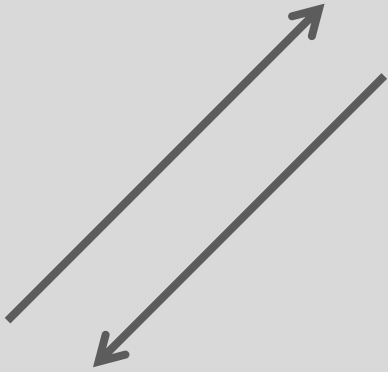


## 04. 코사인 유사도

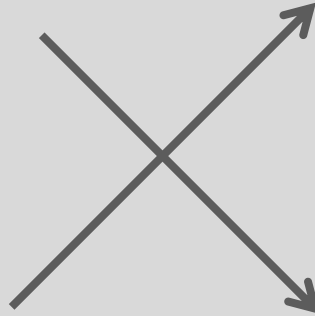
### 코사인 유사도:

두 벡터 간의 사기각을 구하여 유사도 계산

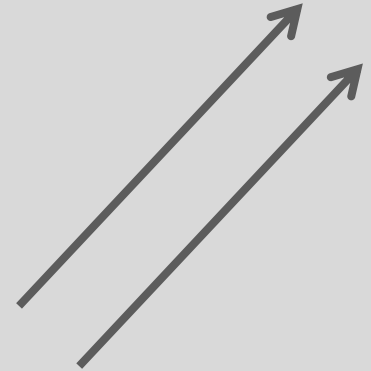
$$\cos(\vec{q} \cdot \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$



코사인 유사도: -1



코사인 유사도: 0



코사인 유사도: +1

## 05. 워드 임베딩

Input :: Embedding vectors

전처리와 TextRank 과정을 통해  
요약된 질문, 답변 데이터셋을 토큰나이징

훈련 데이터의 질문, 답변 데이터셋으로 vocab 생성 후  
{단어: 인덱스} 형식의 딕셔너리 완성(word2idx)

텍스트 데이터를 인덱스 딕셔너리를 이용하여 컴퓨  
터가 해석 가능한 숫자벡터로 변환

word2idx

⟨PAD⟩: 1, ⟨SOS⟩: 2, ⟨EOS⟩: 3,  
⟨UNK⟩: 4, '고소' : 5, ...

[ '사기', '고소', '폭력', '가능' ]



[ '15', '5', '50', '312' ]

# 04

## 모델링

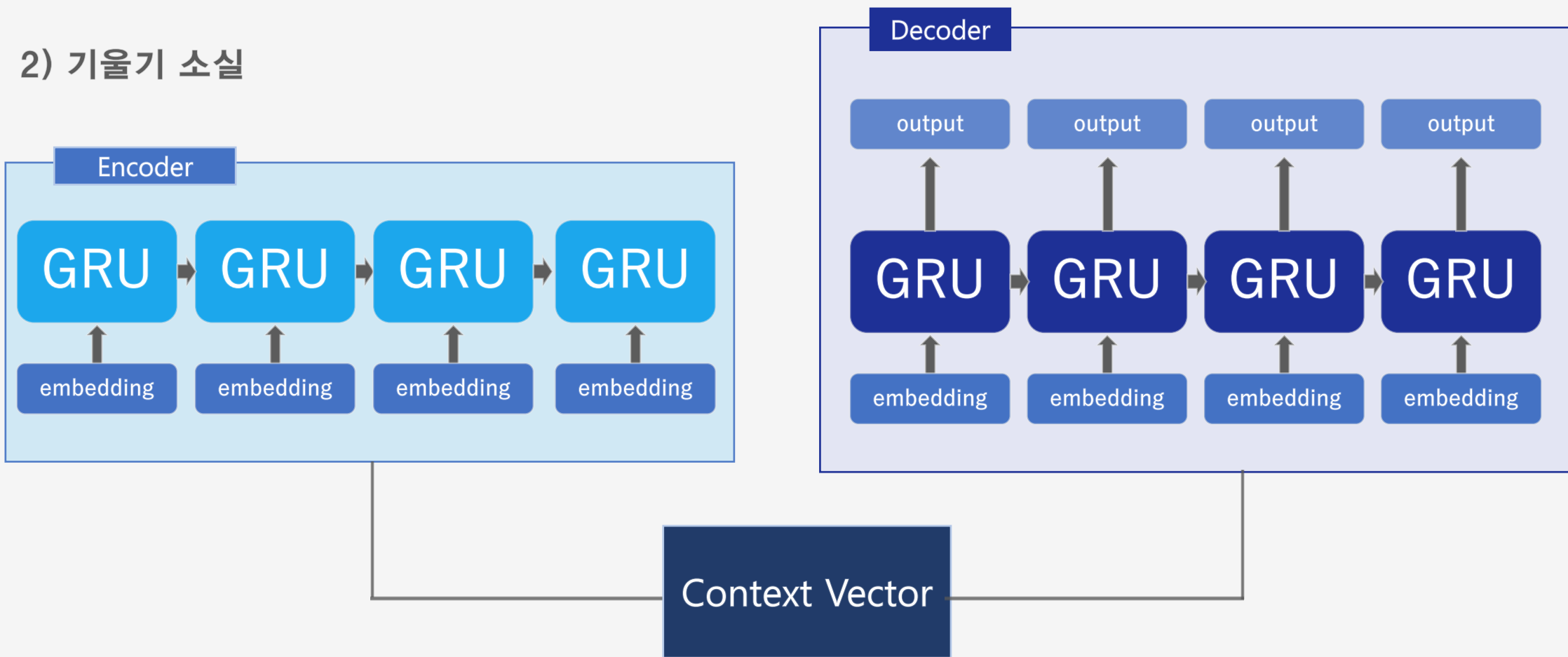
Seq2Seq / Attention / KoGPT2

+

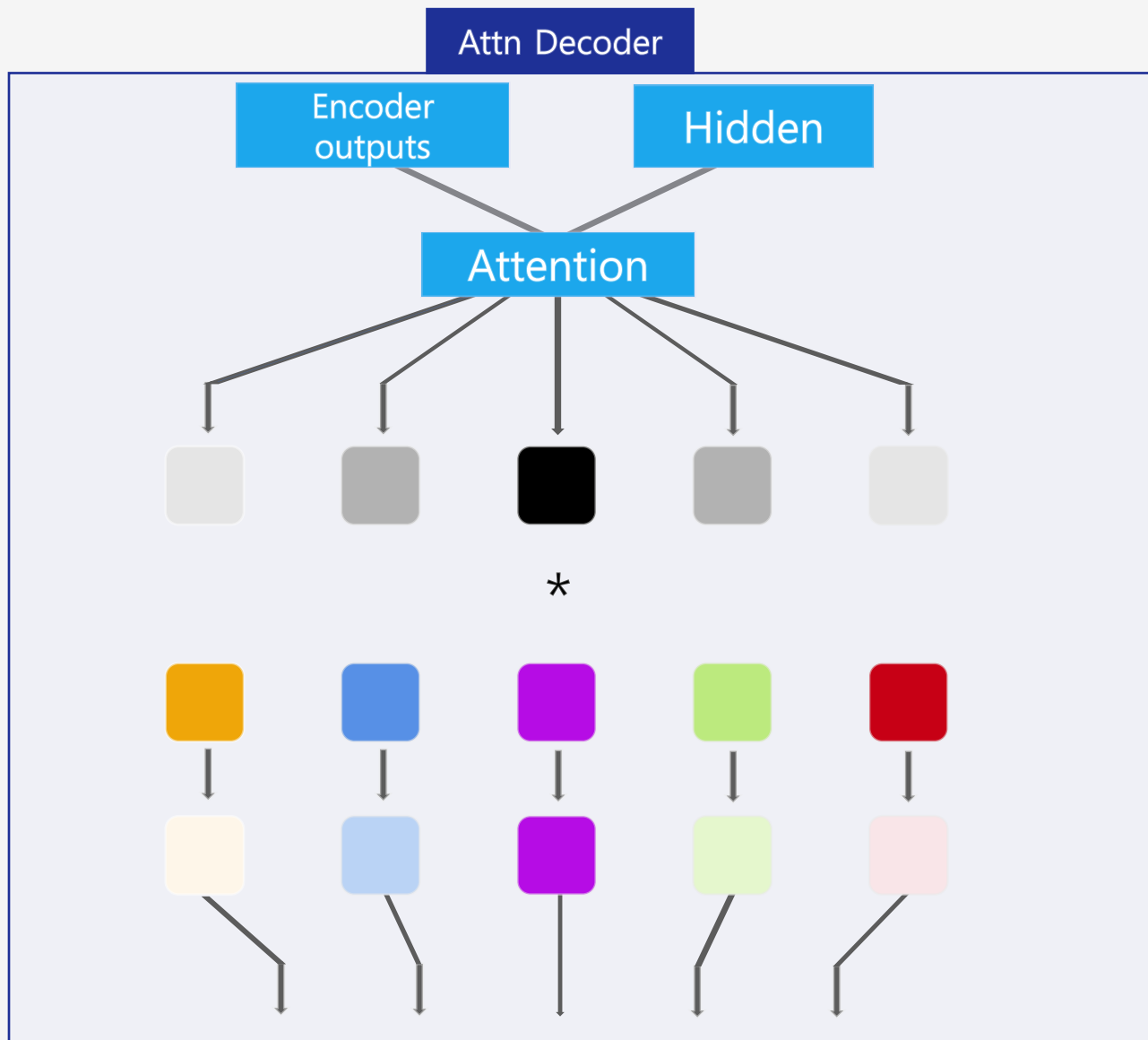
## 06. Seq2Seq

1) 고정된 크기의 Context Vector에  
모든 정보를 압축하기 때문에 정보의 손실 발생

2) 기울기 소실



# 07. Attention



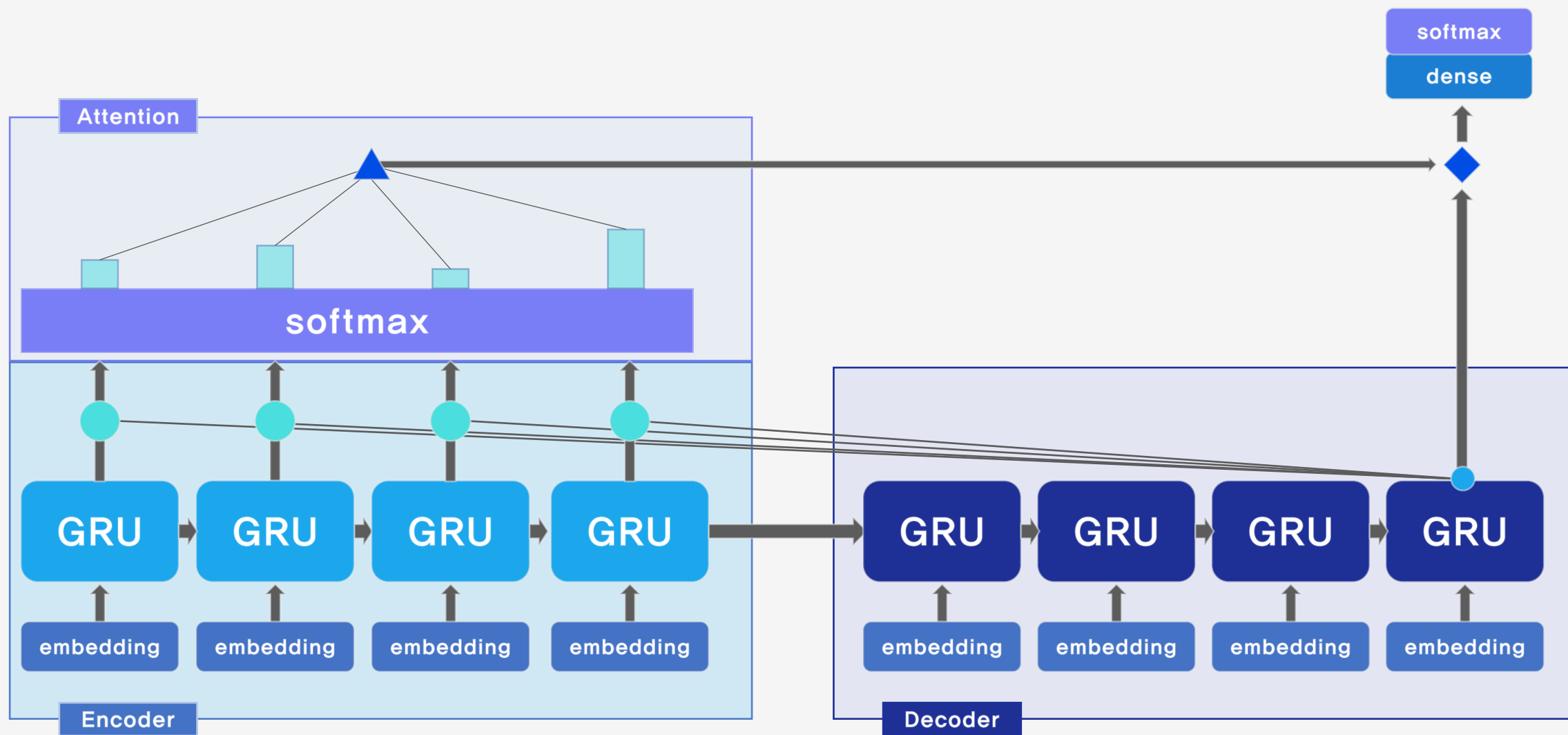
Attention은 출력의  
모든 시점에서 인코더의 출력의  
어텐션 스코어가 큰 부분에 집중

– 정보의 손실을 보완하며  
전체적인 시퀀스 생성

전체 입력 문장을 전부 동일한 비율로 참  
고하는 것이 아니라, 해당 시점에서 예측  
해야 할 단어와 연관이 있는 입력 단어 부  
분에 집중(Attention)

+

## 08. Seq2Seq / Attention



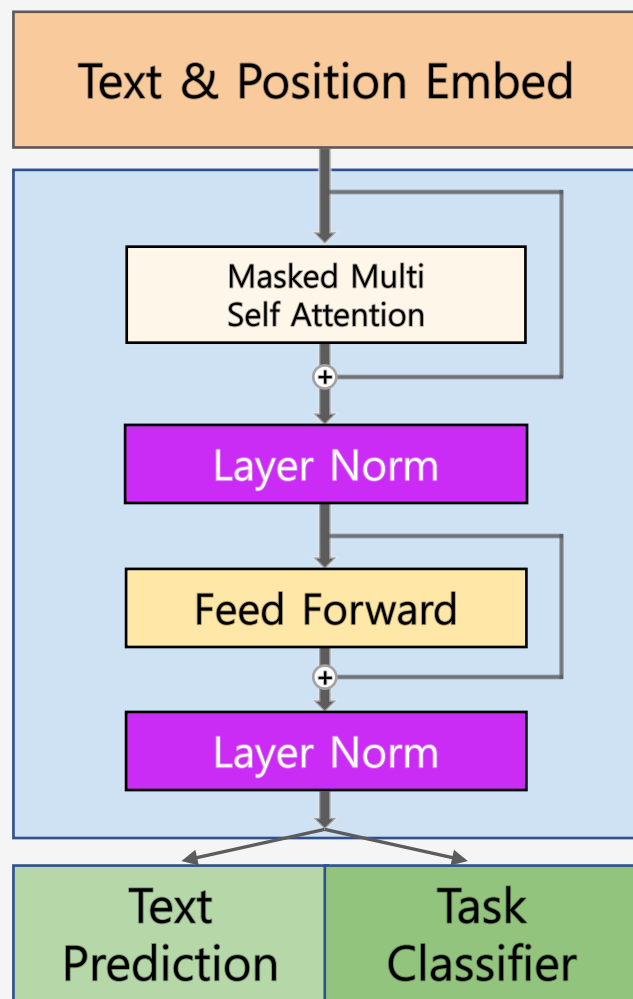
## 09. Seq2Seq 한계

1. Seq2Seq 모델은 RNN 기반 언어 모형이기 때문에,  
긴 문장에 약하고 한정적인 Vocab으로 인한 OOV로 인해  
개인적이거나 지엽적인 사연의 진단이 힘들



Attention 기반 모델이자, 대용량의 말뭉치(Corpus)로  
사전학습 된 **트랜스포머 모델** 을 사용해  
긴 문장에 대한 상담 결과를 반환해보자!

# 10. KoGPT2 Generative Pre-Training



– 트랜스포머의 출력 부분을 처리하는 디코더 부분만 사용

– 앞에서 뒤로 데이터를 순차적으로 처리해 자연어 생성에 유리

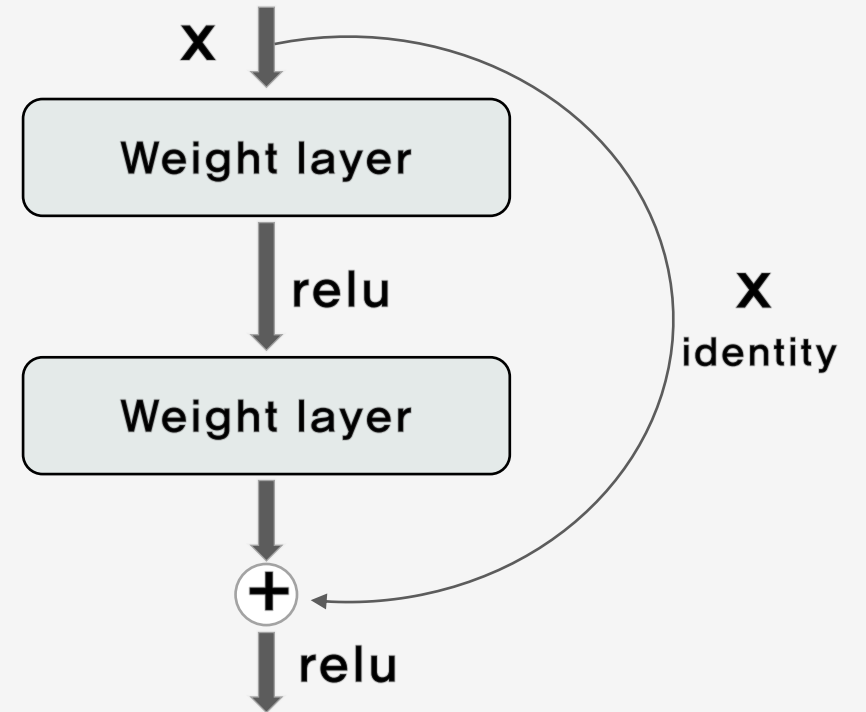
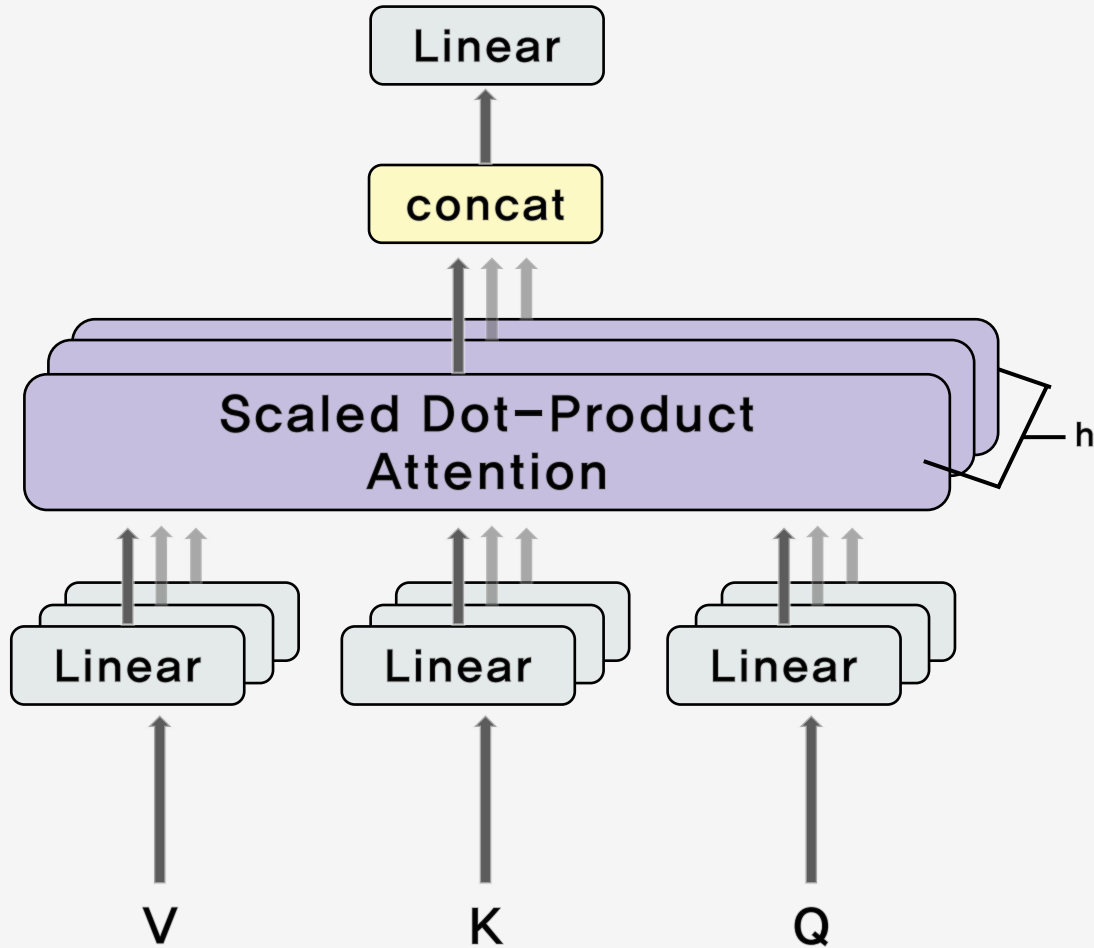
→ 질문과 답변의 쌍으로 전이 학습을 진행해, 질문이 들어왔을 때, 다음에 나타날 가장 높은 확률의 토큰을 예측!



+

# 10. KoGPT2

Attention, Scaled Dot-Product, Multi-Headed  
Attention, Residual Block



# 11. 답변 후처리

## 1. 공백제거, Hanspell 맞춤법 검사기

다소 다툼 여지 가 있 어 보이 는 사안 입니다 다만 임대차법 은 법원 에서 도 유심히 살피 고 있 는 최  
신개정법률 이므로 나중에 되 는 대 로 주장 하기 보다는 사전에 갱신거절 사유 를 명확히 정리 한 뒤  
소송 등 절차 에 돌입 하는 것 이 좋습니다

공백제거

다소다툼여지가있어보이는사안입니다다만임대차법은법원에서유심히살피고있 는최신개정법률이므  
로나중에되는대로주장하기보다는사전에갱신거절사유를명확히정리한뒤소송등절차에돌입하는것이좋습  
니다

Hanspell

다소 다툼 여지가 있어 보이는 사안입니다 다만 임대차법은 법원에서 유심히 살피고 있는 최신개정법  
률이므로 나중에 되는 대로 주장하기보다는 사전에 갱신거절 사유를 명확히 정리한 뒤 소송 등 절차에  
돌입하는 것이 좋습니다

# 11. 답변 후처리

## 2. 인사말추가

### Before

임대차법은 법원에서도 유심히 살피고 있는 최신개정법률이므로 나중에 되는 대로 주장하기보다는 사전에 갹신거절 사유를 명확히 정리한 뒤 소송 등 절차에 돌입하는 것이 좋습니다

인사말 추가



### After

안녕하세요. 법률 상담 AI 서비스입니다.  
질문에 대한 답변드리겠습니다.

임대차법은 법원에서도 유심히 살피고 있는 최신개정법률이므로 나중에 되는 대로 주장하기보다는 사전에 갹신거절 사유를 명확히 정리한 뒤 소송 등 절차에 돌입하는 것이 좋습니다

질문에 대한 충분한 대답이 됐길 바라며,  
자세한 내용은 변호사와 대면하여 진행하시길 바랍니다.

## 12. 답변 예시

### Question

Q : 지급명령을 통해 채권액은 모두 받았습니다.  
 이의신청 없이 도달한지 주 만에 전액을 받을 수 있었습니다.  
 뭐 실제 지급받은 원금에 비하면 송달료 등은 워낙 작은 금액이라  
 거기까지 받을 생각은 없습니다. 이후 채무자가 전화 와서 모든 금액을  
 다 변제했으니 관련된 소송을 취하해달라고 합니다.  
 채권액을 모두 변제받은 현재 혹시 제가 추가로 뭔가를 해야 하는지 궁금합니다

### Answer

A: 안녕하세요. 법률 상담 AI 서비스입니다. 질문에 대한 답변드리겠습니다.  
 기타 채권추심에 필요한 집행권원 및 강제집행 방법에 대해서는 아래 내용을 참고하시  
 기 바랍니다 채권추심에 필요한 집행권원 및 강제집행 방법에 대해서는 아래 내용을 참  
 고하시기 바랍니다 그래야 채권추심 절차를 시작할 수 있습니다 부동산 강제집행 채권추  
 심에 있어서 가장 확실하고 효과적인 방법은 부동산에 대한 강제집행 즉 경매를 신청하  
 는 것입니다 다만 이를 위해서는 채권을 특정해야 하는 데 치밀한 사전조사가 필요합니  
 다 재산 명시 신청 및 재산조회 채권자는 법원에 채무자로 하여금 강제집행의 대상이 되  
 는 재산과 그 재산의 일정한 기간 내의 처분 상황을 명시한 재산목록을 제출하도록 명령  
 을 내려달라는 재산 명시 신청을 할 수 있습니다.  
 질문에 대한 충분한 대답이 됐길 바라며, 자세한 내용은 변호사와 대면하여 진행하시길  
 바랍니다.

## 12. 답변 예시

### Question

Q : 임대인 입니다. 계약 갱신 청구권 요청한 세입자를 내보내려고 하는데 방법이 없을까요  
그러자 세입자는 저에게 내용증명서 형식으로 계약 갱신 청구권을 보내왔습니다.  
검색해보니 임대인 본인이 직접 들어갈꺼라고 하면 대부분 다 해결된다고 나와있더군요  
지식인에도 검색해보니 임대인 본인이 들어갈꺼라고 주장하면  
세입자들은 다 나갈꺼라고 나와있구요.  
그러자 세입자는 임대인이 먼저 매매를 위해 나가라고했고  
그래서 계약 갱신 청구권을 내용증명서 형식으로 보낸것이고  
그로인해 갱신은 이미 완료가 되었으니 나는 나갈 이유가 없고 나갈 생각도 없다고 주장합니다

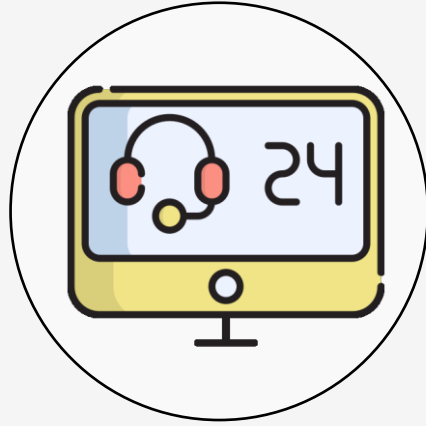
### Answer

A: 안녕하세요. 법률 상담 AI 서비스입니다. 질문에 대한 답변드리겠습니다.  
다소 다툼 여지가 있어 보이는 사안입니다 다만 임대차법은 법원에서도 유심히 살피고  
있는 최신개정법률이므로 나중에 되는 대로 주장하기보다는 사전에 갱신거절 사유를 명  
확히 정리한 뒤 소송 등 절차에 돌입하는 것이 좋습니다 질문 내용만으로는 직접 거주하  
겠다는 내용 외에는 다른 내용이 엿보이지 않는데 다른 사유에도 해당되는 게 있다면 미  
리 정리하여 주장할 필요가 있습니다 앞서 말씀드린 바와 같이 결국 상대방의 항변도 이  
에 집중될 것이므로 사전 준비작업에 유의하시기 바랍니다 실거주 외에도 다른 사유를  
통하여 가능성을 높여야 합니다.  
질문에 대한 충분한 대답이 됐길 바라며, 자세한 내용은 변호사와 대면하여 진행하시길  
바랍니다.

# 04

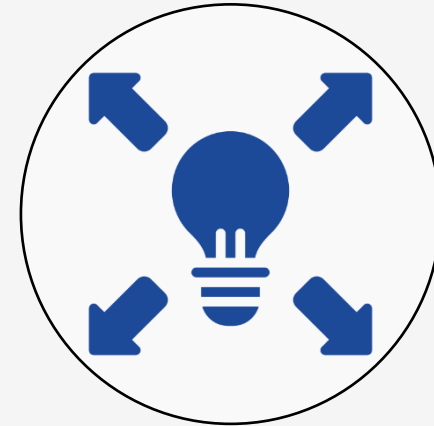
## 기대효과 및 한계점

# 01. 기대효과



## 높은 접근성

법률적인 문제가 생겼을 때,  
부담스러운 변호사와의 직접 상담보다  
간편하게 법률적인 조언 상담 가능



## 확장 가능성

현재의 제한적인 카테고리보다  
더 많은 데이터 수집을 통해,  
상담 내용의 정확성을 높이고  
다양한 법적 카테고리에 대한 상담 가능

## 02. 후기 및 한계점



### 데이터 부족

컴퓨팅 자원과 시간적인 한계로  
데이터 수집량 제한

만약 대용량 데이터를 수집할 수 있으면,  
트랜스포머 계열의 고도의 모델 학습 가능



### 자원 부족

KoGPT 하드웨어 사양  
최소 VRAM 32GB 등등...



---

05

## 시연 및 Q&A

Flask 시연

---

발표를 들어주셔서  
감사합니다 :) )

