

멀티캠퍼스_2팀_자연어처리

자동채점 프로그램

Ordinary

천승진 황도연 권혁재
백혜수 주혜린

CONTENTS

- 기획배경
 - ➔ 문제점 인식
 - ➔ 주제선정
 - ➔ 타깃설정

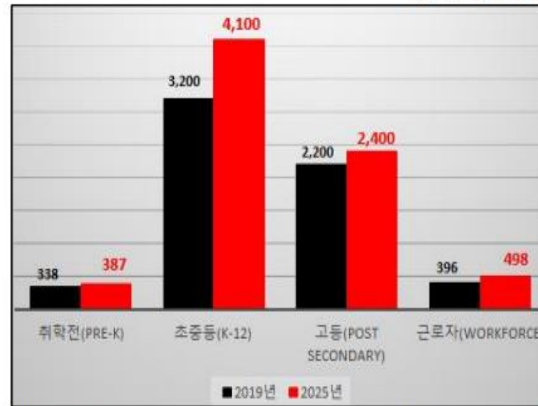
- 자연어
 - ➔ Word2vec
 - ➔ BERT

- 웹구현
 - ➔ 로고 & 챗봇
 - ➔ 시연 동영상



그림 1 글로벌 교육 시장 규모(2019년~2025년)

(단위: USD Billions)



자료: <https://www.holoniq.com/notes/sizing-the-global-edtech-market>, 재구성

그림 2 글로벌 에듀테크 시장 규모(2019년~2025년)



자료: <https://www.holoniq.com/notes/sizing-the-global-edtech-market>, 재구성

학습 도구로서의 서술형 평가 그리고 인공지능의 활용: WA3I 프로젝트 사례

Assessment as a learning tool and utilization of artificial intelligence: WA3I project case

AI 기반 교육 현황과 기술 동향 [PDF 원문보기](#) [인용](#)

Survey of Recent Research in Education based on Artificial Intelligence

쓰기 자동채점 알고리즘의 성능 비교_최대 엔트로피 기법과 서포트 벡터 회귀 기법

The comparison of performances of the automated scoring algorithms: the maximum entropy and support vector regression methods

‘수능 절대평가 전환’ ‘AI 채점 논·서술형 문항 도입’...교육시민단체 고교학점제 안착 제안

2021.12.06 17:50
이호준 기자

中, 시험 채점에 AI 도입... 단순 암기 방지 vs 사회적 파장 논란

오수현 기자 | 입력 2018-05-29 00:01

✓ 교육 시장의 규모 지속적인 증가

✓ 다수의 논문과 기사를 통해 할 수 있는 자동 채점에 대한 사회적 관심과 우려

현재 변별력을 위해 서술형 문제 도입은 필요하지만, 공정성 확보와 사교육 확대라는 사회적 이슈에 대한 대비가 필요.

[교육칼럼] 2028년부터 미래형 수능 도입, 서/논술형 수능 검토한다는데...

수능 서술형 문항 도입에 따른 인프라는 확보된 상태다. 2017년 한국교육과정평가원은 '한국어 서답형(주관식) 문항 자동채점 프로그램'을 개발해 특허를 취득했다. 평가원 관계자는 2016년 국가수준 학업성취도평가 표집 채점에 적용한 결과, 정확도가 100%에 가깝고 프로그램을 활용하지 않고 직접 채점했을 때보다 높은 수준이라고 밝혔다.

출처 : 조선에듀 2021.07.29

But..

오락가락하는 AI 선생님...채점 기준 제각각

온라인 교육 플랫폼 캔버스, 오투스 문제점
대문자로 썼다고 오답, 단위 표기했다고 오답
업체 측 “정답 설계하는 담당교사들 책임회피

출처 : '인공지능 전문미디어' AI타임스 2020.11.19

현업자(교수) 인터뷰

AI가 대신 채점해 주면 사람이 하는 것보다 더 정확할 것이다. 하지만, 사회적으로 용인될 수 있는 부분인지는 따져봐야 한다. 인간이 하지 않는 만큼, 진입 장벽은 높고 작은 실수도 용납되지 않을 것이다.



영어 주관식 서술형 자동채점 프로그램

채점의 끝이라는 뜻으로, 영어 서술형 답안을 자동으로 채점해 주는 AI서비스

- 다량의 데이터를 높은 정확도로 자동 채점해주는 기능
- 자동 채점에 사용된 데이터 및 결과물을 저장하고, 저장한 데이터를 찾을 수 있는 DB기능



사교육



수능 및 학교 시험



자격증 및 기업 시험



다루는 과목 : 사회/과학 부문의 영어

채점 대상 : 중/고등학생

채점 기준 : 키워드 매칭, 문맥적 의미 유사도, 문법 정확도

최종 목표 :

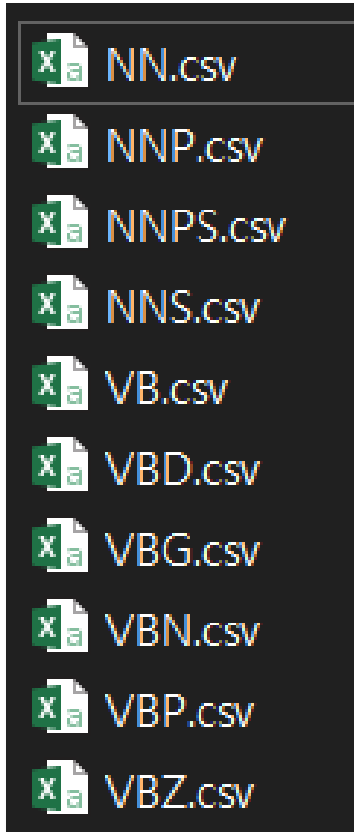
- 키워드 기준, 문맥적 의미 기준, 문법적 정확도 기준에 대한 학생 점수 제공.
- 학생들 답안 csv 안에 가채점 점수가 표기된다.



진행일정

2022년 8월							< 오늘 >	
일	월	화	수	목	금	토		
31	8월 1일	2	3	4	5	6		
7	8	9	10	11	12	13		
14	15	16	17	18	19	20	아이데이션..	
21	22	23	24	25	26	27	아이데이션 & 멘토링	
자연어처리 개념 및 전처리 스터디								
데이터 취득 및 전처리								
28	29	30	31	9월 1일	2	3	자연어처리 개념 및 전처리 스터디	
데이터 취득 및 전처리								
BERT, GPT-3 등 딥러닝 기법 스터디								

2022년 9월							< 오늘 >	
일	월	화	수	목	금	토		
28	29	30	31	9월 1일	2	3	자연어처리 개념 및 전처리 스터디	
데이터 취득 및 전처리								
BERT, GPT-3 등 딥러닝 기법 스터디								
4	5	6	7	8	9	10	데이터 취득 및 전처리	
BERT, GPT-3 등 딥러..							※ 추석	
모델링 진행								
11	12	13	14	15	16	17	모델링 진행	
※ 추석								
웹개발 스터디								
18	19	20	21	22	23	24	모델링 진..	
웹 개발							웹개발 ..	
25	26	27	28	29	30	10월 1일	웹 개발	
PPT 제작 및 발표준비								



VB	word	type	v1
abolish			
abolitionize			
absolutize			
academize			
acclimatize			
acculturize			
acetalize			
acetize			
acetoneize			
acetylyze			
achromatize			
acquire			
actify			
actionize			
activize			
actualize			
adonize			
aegagropila			
agatize			
agglutinate			
agnize			
aitch			



Column1	word	type	v1	v2	v3	유의어	v4	v5	반의어	v6	동의어
0	abolish	VB	(법률·제도·조직을) 폐지하다	(법률·제도·조직을) 폐지하다	(법률·제도·조직을) 폐지하다		(법률·제도·조직을) 폐지하다		approve	폐지하다	annul
1	acclimatize	VB	(장소·기후 등에) 익숙해지다(적응하다)	(장소·기후 등에) 익숙해지다(적응하다)	(장소·기후 등에) 익숙해지다(적응하다)		(장소·기후 등에) 익숙해지다		refuse	익숙해지다	accommodate
2	actualize	VB	현실로 만들다, 실현하다	현실로 만들다, 실현하다	현실로 만들다, 실현하다		현실로 만들다, 실현하다			현실로 만들다	engineer
3	algebraize	VB	대수의, 대수적인	대수의, 대수적인	대수의, 대수적인		대수의, 대수적인			대수의, 대수적인	NaN
4	alphabetize	VB	알파벳순으로 배열하다	알파벳순으로 배열하다	알파벳순으로 배열하다		알파벳순으로 배열하다		disorganize	알파벳순으로 배열하다	order
5	amortize	VB	[상업] (빚을) 분할 상환하다	(빚을) 분할 상환하다	(빚을) 분할 상환하다		(빚을) 분할 상환하다			분할 상환하다	NaN
6	antagonize	VB	적대감을 불러일으키다	적대감을 불러일으키다	적대감을 불러일으키다		적대감을 불러일으키다		calm	적대감을 불러일으키다	anger
7	anthologize	VB	(작가나 작품을) 선집에 포함시키다	(작가나 작품을) 선집에 포함시키다	(작가나 작품을) 선집에 포함시키다		(작가나 작품을) 선집에 포함시키다		disarrange	선집에 포함시키다	arrange
8	apprize	VB	[격식] 알리다	알리다	알리다		알리다		hide	알리다	brief
9	asexualize	VB	무성화하다, 생식 능력을 없애다	무성화하다, 생식 능력을 없애다	무성화하다, 생식 능력을 없애다		무성화하다, 생식 능력을 없애다		assist	무성화하다	neuter
10	atomize	VB	원자화하다, 세분화하다	원자화하다, 세분화하다	원자화하다, 세분화하다		원자화하다, 세분화하다		construct	원자화하다	grate
11	baconize	VB	베이컨을 첨가하다, 베이컨 맛을 내	베이컨을 첨가하다, 베이컨 맛을 내	베이컨을 첨가하다, 베이컨 맛을 내		베이컨을 첨가하다, 베이컨 맛을 내			베이컨을 첨가하다, 베이컨 맛을 내	NaN
12	bipolarize	VB	[심리] 조울증의, 조울증을 앓는	조울증의, 조울증을 앓는	조울증의, 조울증을 앓는		조울증의, 조울증을 앓는			조울증의, 조울증을 앓는	NaN
13	blemish	VB	[격식] 흠집을 내다	흠집을 내다	흠집을 내다		흠집을 내다		perfection	흠집을 내다	blotch
14	canalize	VB	[전문 용어] (강을) 운하로 만들다	(강을) 운하로 만들다	(강을) 운하로 만들다	2	(강을) 운하로 만들다	2		운하로 만들다	NaN
15	carburize	VB	[전문 용어] 탄소를 혼합하다	탄소를 혼합하다	탄소를 혼합하다		탄소를 혼합하다			탄소를 혼합하다	NaN
16	categorize	VB	분류하다 (=classify)	분류하다 (=classify)	분류하다 (=classify)		분류하다 (=classify)			분류하다	pigeonhole
17	centralize	VB	중앙집권화하다	중앙집권화하다	중앙집권화하다		중앙집권화하다		separate	중앙집권화하다	integrate
18	centrifugal	VB	(=centrifuge)	(=centrifuge)	(=centrifuge)		(=centrifuge)				NaN
19	civilize	VB	개화(교화)하다, (태도를) 세련되게 하다	개화(교화)하다, (태도를) 세련되게 하다	개화(교화)하다, (태도를) 세련되게 하다		개화(교화)하다, (태도를) 세련되게 하다		condemn	개화하다, (태도를) 세련되게 하다	enlighten
20	clarify	VB	[격식] 명확하게 하다, 분명히 말하다	명확하게 하다, 분명히 말하다	명확하게 하다, 분명히 말하다	2	명확하게 하다, 분명히 말하다		confuse	명확하게 하다, 분명히 말하다	clear up
21	codify	VB	[전문 용어] (법률 등을) 성문화하다	(법률 등을) 성문화하다	(법률 등을) 성문화하다		(법률 등을) 성문화하다		disorder	성문화하다	arrange

- 명사, 형용사, 동사를 중심으로 데이터셋 구성.
- 해당 단어들 크롤링하여 사전 구성.
- 유의어 and 반의어 포함 설정.

문제

키워드 & 답안 별 키워드 존재 유무

	7. The roots of plants are underground, so they are hard to see, but they have many functions. What do plant roots do?	q7_storage	q7_support	q7_carry	q7_storage	q7_support	q7_Transport	1
0	It absorbs water that sprays plants							
1	water absorption							
2	Absorption of nutrients, absorption							
3	water absorption							
4	It absorbs water							
5	It absorbs water and nutrients through its roots to help plants grow well							
6	It absorbs water from the ground and makes it available. It often stores nutrients.							
7	Plant fixation, nutrient absorption							
8	The roots absorb water as much as possible by widening the surface area as possible.							
9	water absorption							
10	Plant roots absorb nutrients and water from the ground and carry them to stems and leaves.							
11	It functions to absorb nutrients.							
12	Filter using the intubation pressure function and transport energy to the body and water pipes.							

학생답안

OMR 서술형 답안

1	The liquid is less reactive
2	Day and night change due to the rotation of the Earth
3	It will move to where the battlefield flows
4	The air in the bag would have turned into air that was not fresh
5	I can breathe



OCR 인식한 결과

	stu_ans					
0	The liquid is less reactive					
1	Day and night change due to the rotation of the Earth					
2	It will move to where the battlefield flows					
3	The air in the bag would have turned into air that was not fresh					
4	I can breathe					



최종 데이터셋

En_right_answer 문제 별 모범 답안지

question	right_ans
q1	Since the circuits are connected in parallel, each electronic product is independently supplied with power. Due to the fast power cut off due to the operation of the air conditioner, the power of the air conditioner is quickly cut off.
q1	Since the circuits are connected in parallel, each electronic product is independently supplied with power. The power of the air conditioner is quickly cut off due to the operation of the air conditioner.
q2	Due to convection, dense air sinks due to its large density, and warm air rises due to its small density. Due to convection, it is efficient to install warm air below the ceiling.
q2	Due to convection, cold air sinks due to its large density, and warm air rises due to its small density. Therefore, it is efficient to install warm air below the ceiling.
q3	Since the volume of the liquid is smaller than the volume of the gas, storing gaseous fuel in a liquid state can store a large capacity in a small volume and can be stored safely.
q3	Since the volume of the liquid is smaller than the volume of the gas, if the gaseous fuel is stored in a liquid state, a large capacity can be stored in a small volume and can be stored safely.
q4	When oxygen is lost, all combustion and oxidation reactions disappear, and all life phenomena disappear because a large proportion of elements disappear.
q4	When oxygen is lost, all combustion and oxidation reactions disappear, and all life phenomena disappear because a large proportion of elements disappear.
q5	When sweat is released, water in the body disappears, so it lowers body temperature to maintain homeostasis. And the sweat is vaporized and absorbs the heat of the body.
q5	When sweat is released, moisture in the body disappears, so it lowers body temperature and maintains homeostasis. And the sweat released vaporizes to absorb the heat of the body.
q6	Each country has different crops grown, and they are exported from more produced countries because they can be sold cheaper in that country. Therefore, the crops grown in each country are different.
q6	The crops grown in each country are different. Countries that produce more crops export them because they can be sold cheaper in that country. The crops grown in each country are different.
q7	The roots of plants carry moisture in the ground to the stems and leaves and maintain life, store nutrients, and support plants deeply embedded in the ground.
q7	The roots of plants carry moisture in the ground to the stems and leaves and maintain life, store nutrients, and support plants deeply embedded in the ground.
q8	Because of the simple food chain, an ecosystem with low biodiversity does not have a significant impact on the ecosystem because there are many organisms that feed on them.
q8	Ecosystems with low biodiversity have simple food chains, so if any species suddenly disappears, other organisms that feed on them can also disappear.
q9	Water evaporates and becomes water vapor due to the rising air current, and the water vapor expands adiabatically and condenses to form a cloud.
q9	Water evaporates due to rising air currents and becomes water vapor. And the water vapor expands adiabatically and condenses to form a cloud.

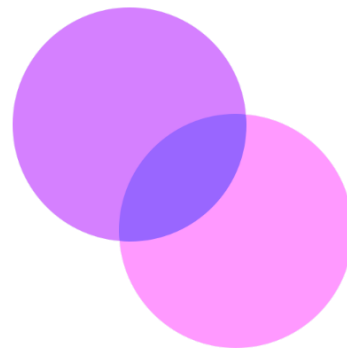
Student_key_q1~64.csv -> 64개
문제와 학생 답, 그리고 keyword가 들어있는 데이터셋
총 64문제이므로 64개의 데이터셋

question	right_ans	keyword
3. The butane gas fuel used to cook food outdoors is liquid in the butane gas cylinder. Why is it liquid?	q3_particle arrangement q3_storage q3_reactivity q3_confirmation	
0 I don't know.		
1 It's very responsive		
2 Low boiling point and remains gaseous at room temperature		
3 To make the volume smaller		
4 This is because if stored in a gaseous state, all of them will be blown away.		
5 This is because the boiling point of butane gas is low at room temperature		
6 I don't understand.		
7 Low boiling point and cannot be stored as liquid at room temperature		
8 This is because the heat of vaporization generated when a liquid becomes a gas absorbs ambient heat.		
9 for storage		
10 Butane in a gaseous state is stored in a liquid state with low reactivity because it has high reactivity.		
11 At the same amount, the volume of the gas is larger than the volume of the liquid, so it must be stored as a liquid to store it in less space.		
12 Since it is stored in a high pressure state, a combustible gas such as butane gas can be used relatively safely.		
13 This is because gases react with other substances.		
14 As the pressure is lowered, the gas maintained in a liquid state by the high pressure expands again and becomes gas.		
15 It exists as a liquid because of the pressure of pressing butane gas, but if it exists as a liquid, the distance between molecules and molecules is very short.		
16 Gas is bulky and difficult to carry around, so it is carried in a small liquid state		
17 It exploded		
18 This is because it needs to be liquid to prevent deterioration as much as possible.		
19 The liquid is less reactive.		
20 To prevent it from expanding and bursting.		
21 Since the gas is larger in volume than the liquid, when stored in a gaseous state, it is stored in a smaller amount than when stored in a liquid state.		



Keyword Checker

- 학생 답에 해당 키워드가 있으면 1
- 해당 키워드가 없으면 0



SBert

- 문장 유사도 검출



Language_tool

- 문법 정확도 파악

Keyword Checker

word2vec

doc2vec

```
from gensim.models import Word2Vec

model = Word2Vec(sentences=commentlist, size=100, window=5, min_count=0, workers=4, sg=0)
model_cleaned = Word2Vec(sentences=commentlist_cleaned, size=100, window=5, min_count=0, workers=4, sg=0)

def get_document_vectors(document_list):
    document_embedding_list = []

    # 각 문서에 대해서
    for line in document_list:
        doc2vec = None
        count = 0
        for word in line.split():
            if word in model_cleaned.wv.vocab:
                count += 1
                # 해당 문서에 있는 모든 단어들의 벡터값을 더한다.
                if doc2vec is None:
                    doc2vec = model_cleaned[word]
                else:
                    doc2vec = doc2vec + model_cleaned[word]

        if doc2vec is not None:
            # 단어 벡터를 모두 더한 벡터의 값을 문서 길이로 나눠준다.
            doc2vec = doc2vec / count
            document_embedding_list.append(doc2vec)

    # 각 문서에 대한 문서 벡터 리스트를 리턴
    return document_embedding_list
```

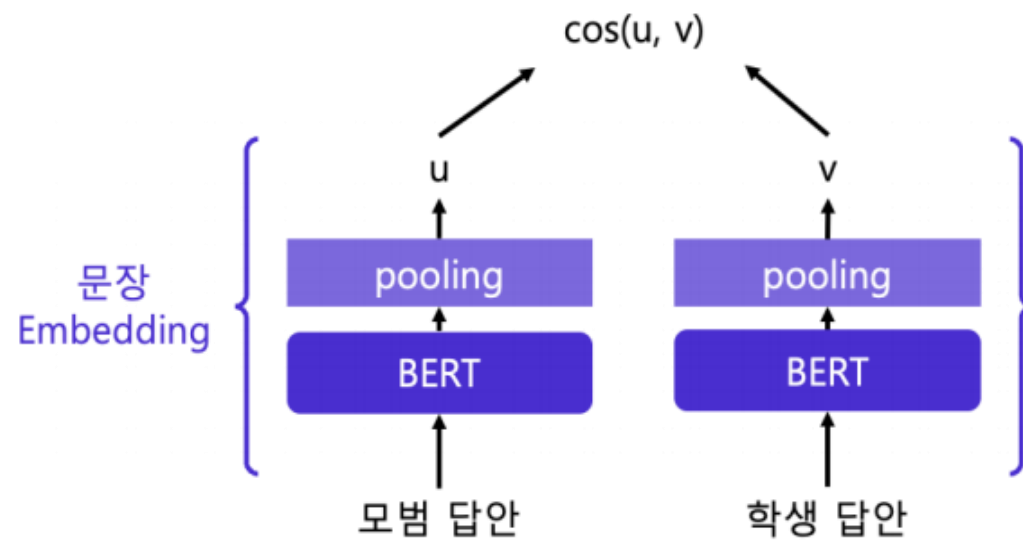


Keyword Checker

Keyword : particle arrangement, storage, reactivity, confirmation

학생 답안	particle arrangement	storage	reactivity	confirmation
for storage	0	1	0	0
Beutane in a gaseous state is stored in a liquid state with low reactivity because it has high reactivity.	0	0	1	0

SBert





Language_tool

```
def grammer_check(filepath):  
  
    tool = language_tool_python.LanguageTool('en-US')  
    for q in filepath:  
        q_pd = pd.read_csv(os.path.join(cos_path,q),index_col=0)  
        # q_pd = q_pd.drop(columns = ['right_ans', 'similarity'])  
        q_pd['check'] = 0  
        print(q_pd)  
        for i in range(len(q_pd)):  
            text = q_pd.iloc[i][1]  
            print(text)  
  
            # get the matches  
            matches = tool.check(text)  
  
            print(matches)  
            print(len(matches))  
            q_pd['check'][i]=len(matches)
```



최종 점수

〈표 2〉 점수 계산을 위한 기준표

점수	형태소오류	구문오류	매핑오류
2	없음	없음	모두 적절
1	1~2개	1개 이하	1~2개 틀림
0	3개 이상	2개 이상	3개 이상 틀림

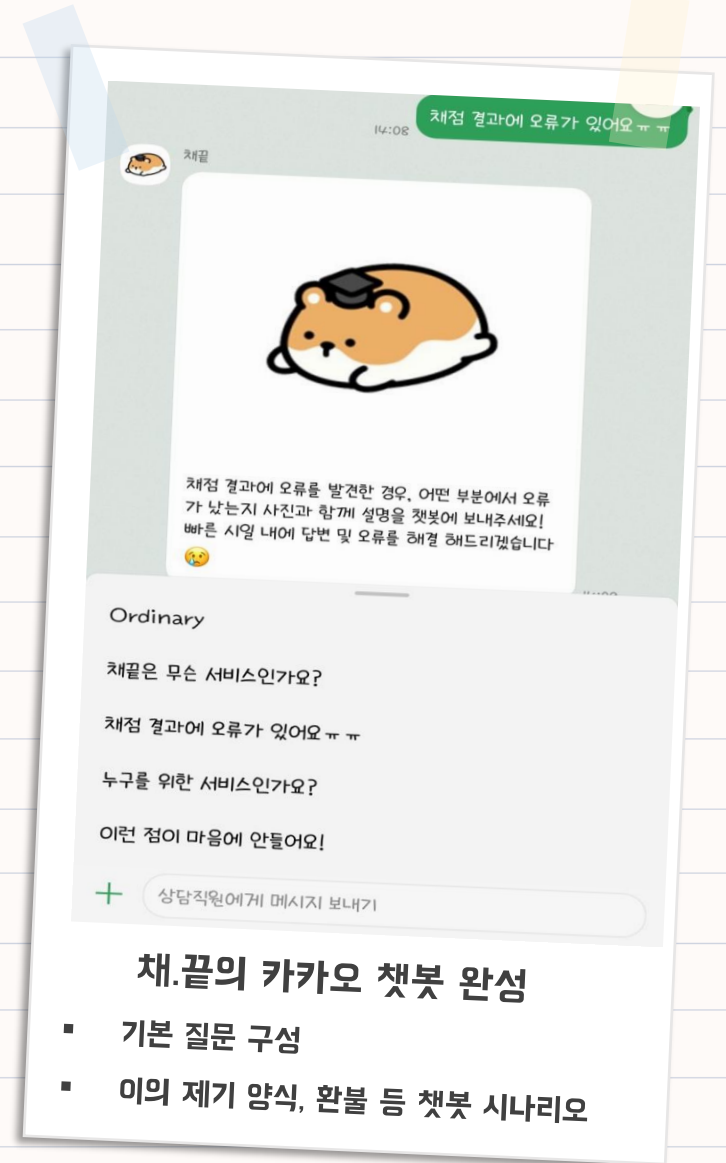
$$\text{문법정확도}(0.6 \times \alpha) + \text{키워드}(0.3 \times \beta) + \text{문장유사도}(0.1 \times \gamma)$$



채.끝의 메인 로고 확정



개인 캐릭터 확정

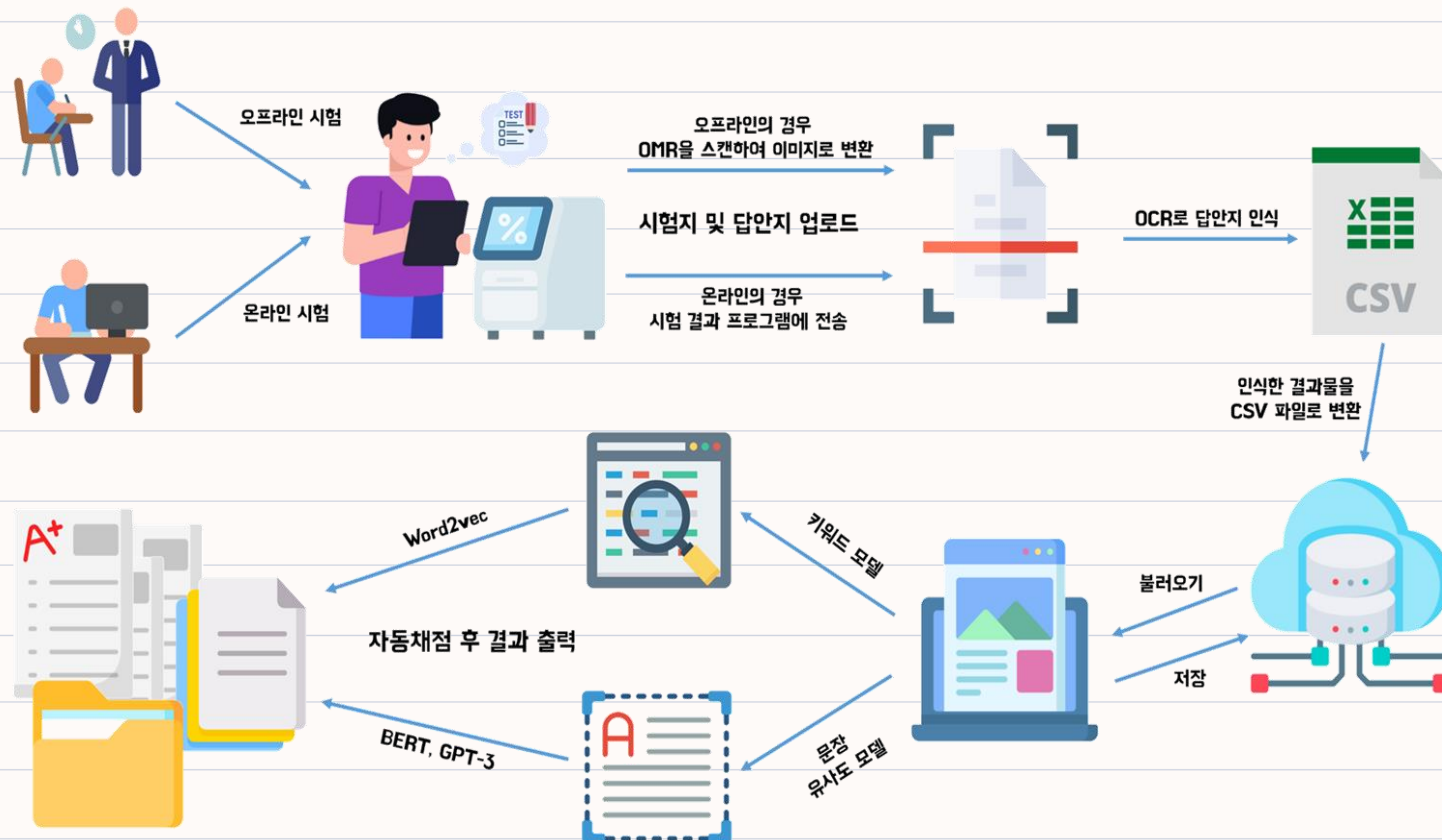


채.끝의 카카오톡 챗봇 완성

- 기본 질문 구성
- 미의 제기 양식, 환불 등 챗봇 시나리오



채점의 끝 시스템 구조



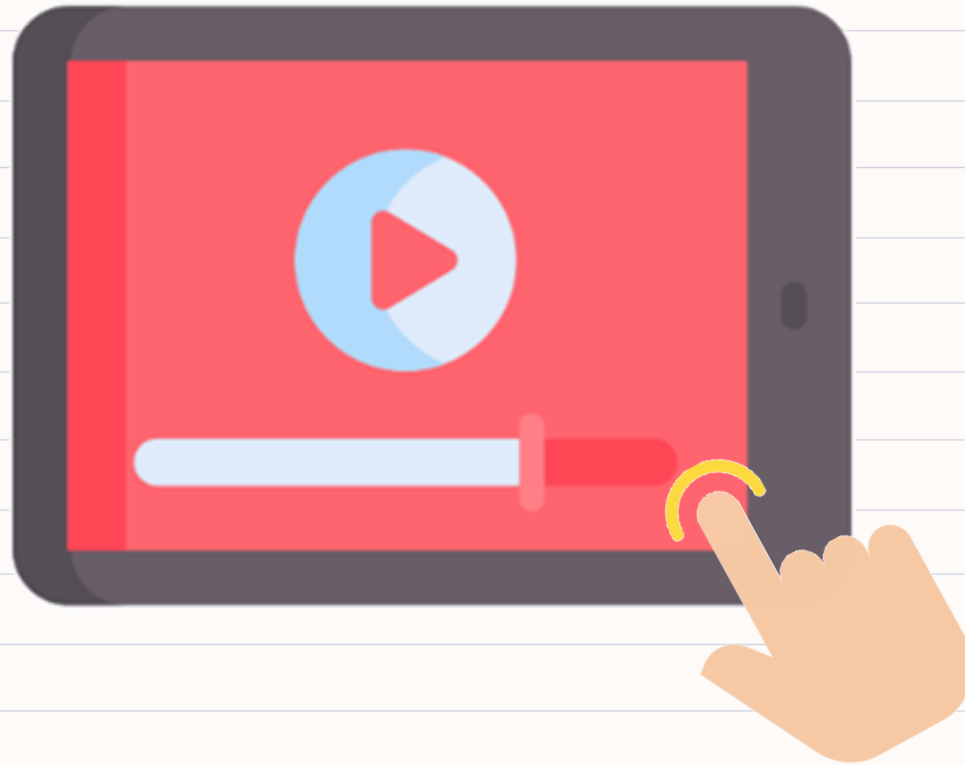
Frontend



Backend

goormide
django





**Thank
you**

