

# 제품 스펙 문서 및 숫자 정보처리 를 위한 표 질의 응답 시스템

201624410 권선근

201624473 민경언

201624542 이상진

# 목차

- 01 연구 배경 및 목표
- 02 연구 배경 지식
- 03 연구 내용
- 04 연구 결과 분석 및 의의

# 연구배경 및 목 표

제품 스펙 문서 질의응답 시스템의 필요성과  
연구목표에 대해 알아보겠습니다

## 연구 배경

1

4차 산업혁명이 대두됨에 따라 머신러닝 기술에 대한 관심이 증가하고 있고, 이것을 활용한 질의응답 시스템이 개발되고 있다.

2

207,000 여개의 위키백과 문서로 학습된 언어모델의 질의응답 기능이 잘 작동하지 않는 경우들을 탐색하였다.

3

기존의 한국어 표에 대한 모델은 제품스펙 표에 대한 자연어 질문처리에 매우 취약한 모습을 보인다.

## 기존 문제점 및 연구 목표

### | 기존 문제점 |



- 언어모델은 위키 문서로만 학습되어 있다.
- 제품스펙 문서의 표와는 유형이 크게 다르다.

### | 연구 목표 |

- 01 제품 스펙 표에 대한 질의에 답을 도출하는 질의응답 시스템 개발
- 02 시스템이 가장 적합한 정답을 도출하도록 파인 튜닝을 진행

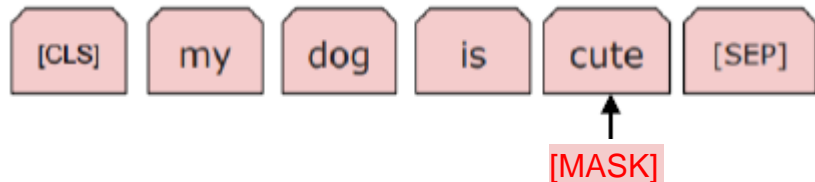
## 연구배경 지식

- 언어모델인 BERT와 이를 확장한 TAPAS에 대한 배경지식을 알아봅니다.

# 02 | 연구 배경 지식

## BERT(Bi-directional Encoder Representations from Transformers)

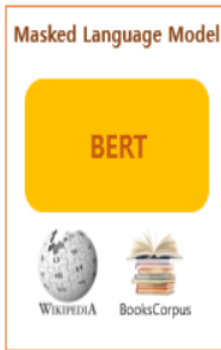
### BERT를 활용한 사전학습



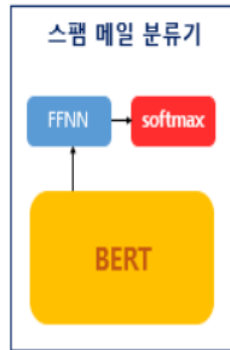
[MASK]를 예측하는 과정에서 성능 증대

- 방대한 양의 데이터를 마스킹 기법을 사용하여 사전학습된 언어 모델

- 특정 단어를 [MASK]로 가리고, 이 위치에 들어가는 기존 단어를 예측하는 과정에서 언어에 대한 지식 학습



조금만 튜닝(Tuning)해서  
다른 용도로 사용한다면?



33억 단어에 대해서 4일간 학습시킨 언어 모델

BERT의 지식을 이용한 스팸 메일 분류기

- 잘 만들어진 bert모델위에 1개의 classification layer만 부착하여 다양한 자연어 처리를 수행.

- 이번 연구과제에서는 스팸 표에 대한 질의응답 데이터를 사용해서 파인튜닝을 수행

## TAPAS(Table Parser)



확장



구글에서 개발한 자연어처리 사전 훈련 언어모델, 표에 대한 질의 응답정확도는 떨어짐

BERT를 기반으로 하며 표에 대한 질의에도 높은 정확도를 보이는 모델

- 이스라엘 텔아비브 대학의 전산언어학회가 2020년에 발표한 기술
- BERT 기반으로 표 데이터에 특화
- 표에 포함된 정보를 자연어로 추출하는 질의응답 모델 기술

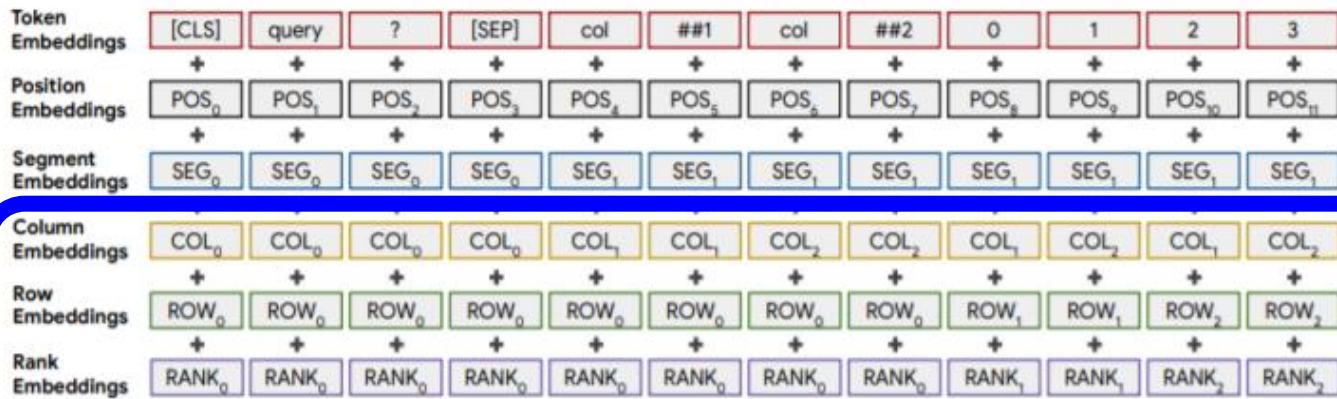


## TAPAS(Table Parser) 6가지 임베딩

Table

col1	col2
0	1
2	3

TAPAS에서 추가



- **Token Embedding:** 입력 텍스트가 split된 토큰에 대한 임베딩
- **Position Embedding:** 토큰의 위치를 가리키는 순서에 대한 임베딩
- **Segment Embedding:** 토큰이 질문 혹은 테이블에 소속되어 있는지에 대한 임베딩

- **Column Embedding:** 토큰의 테이블 상 Column 위치
- **Row Embedding:** 토큰의 테이블 상 Column 위치
- **Rank Embedding:** cell 마다 랭킹을 부여하여 순위를 매기는 특수 순위 인덱스 임베딩

## TAPAS(Table Parser) 의 임베딩 예시

나이가 가장 많은 사람은?



이름	나이
민수	12
철수	15



셀이 평면화되어 들어감

Token Embeddings	[CLS]	...	[SEP]	이름	나이	민수	12	철수	15	[SEP]
Position Embeddings	0	1	2	3	4	5	6	7	8	9
Segment Embeddings	0	0	0	1	1	1	1	1	1	1
Column Embeddings	0	0	0	1	2	1	2	1	2	0
Row Embeddings	0	0	0	1	1	2	2	3	3	0
Rank Embeddings	0	0	0	0	0	0	1	0	2	0

순위를 정할 수 없는 경우 0, 있는 경우  
오름차순 (나이가 많을수록 높은 순위)

## TAPAS 모델의 정답 도출 과정

- TAPAS에서는 질문을 보고 정답의 유형을 Cell Selection, Compute Mode등으로 분류한다.
- Compute Mode 일 경우 질문을 분석하여 가장 확률이 높은 연산자를 사용한다.

Table

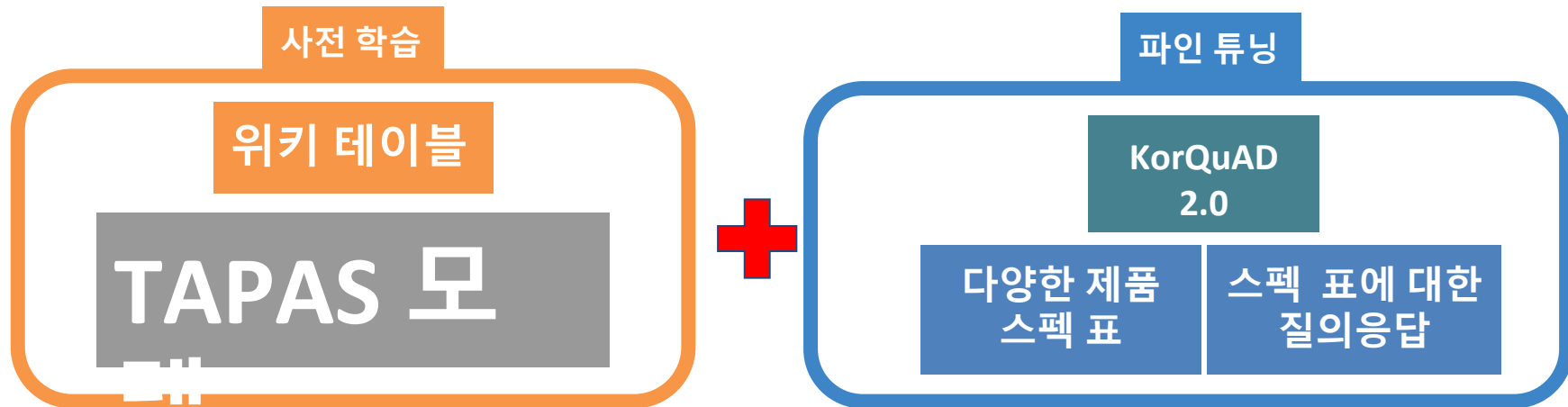
Rank	Name	No. of reigns	Combined days
1	Lou Thesz	3	3,749
2	Ric Flair	8	3,103
3	Harley Race	7	1,799
4	Dory Funk Jr.	1	1,563
5	Dan Severn	2	1,559
6	Gene Kiniski	1	1,131

Example questions

#	Question	Answer	Example Type
1	Which wrestler had the most number of reigns?	Ric Flair	Cell selection
2	Average time as champion for top 2 wrestlers?	AVG(3749,3103)=3426	Scalar answer
3	How many world champions are there with only one reign?	COUNT(Dory Funk Jr., Gene Kiniski)=2	Ambiguous answer
4	What is the number of reigns for Harley Race?	7	
5	Which of the following wrestlers were ranked in the bottom 3?	{Dory Funk Jr., D Severn, Gene Ki	
	Out of these, who had more than one reign?	Dan Severn	

$op$	$compute(op, p_s, T)$
COUNT	$\sum_{c \in T} p_s^{(c)}$
SUM	$\sum_{c \in T} p_s^{(c)} \cdot T[c]$
AVERAGE	$\frac{compute(SUM, p_s, T)}{compute(COUNT, p_s, T)}$

## TAPAS 모델의 학습 과정



- TAPAS를 질의응답 시스템으로 사용하기 위해 위키 테이블로 사전학습
- 한국어 위키피디아 기반의 100,000+쌍으로 구성된 한국어 질의응답 데이터셋인 KorQuAD 2.0 과 직접 수집한 다양한 제품 스펙 표에대한 질의응답 데이터셋을 추가 학습 용도로 사용

## 연구 내용

- 이번 과제에서 수행한 연구 내용에 대해 알아보겠습니다.

## 테이블 데이터 수집 방식

제조사, 쇼핑몰 사이트 등에서 다양한 테이블 데이터를 수집하였다.

### ① 단일 테이블



제조사	QCY (제조사 웹사이트 바로가기)	등록년월	2020년 12월
유/무선 구분	무선	용도	음악+통화용
유닛형태	커널형	블루투스	블루투스 v5.0
무선연결거리	10m		
지원			
HFP	○	HSP	○
AVRCP	○	A2DP	○
부가기능			
핸즈프리통화	○	재생/멈춤조절	○
재생시간	4시간(싱글), 2시간(듀얼)	통화시간	3시간
대기시간	1440시간	충전시간	1시간
배터리용량	43mAh	무게	무게 4.5g

### ② 다중 비교 테이블

이미지			
제조사/브랜드	QCY	삼성전자	APPLE
상품명	T13 (정품)	갤럭시 버즈2 SM-R177 (정품)	에어팟 프로 (정품)
희재가격	17,970원	126,510원	230,820원
등록년월	2021.06	2021.08	2019.10
유/무선 구분	무선	무선	무선
용도	음악+통화용	음악+통화용	음악+통화용
유닛형태	커널형	커널형	커널형
유닛크기	유닛: 7.2mm		
블루투스	블루투스 v5.1		블루투스 v5.0
지원	SBC AAC HFP HSP AVRCP A2DP	SBC AAC	AAC

## 테이블 데이터 전처리

### 테이블 데이터 전처리 전후 비교

이미지			
	39.6cm / 2.25kg 라이젠5 / RTX3050	39.6cm / 1.69kg 11세대 / MX450	
제조사/브랜드	레노버 게이밍	레노버 아이디어패드	ASUS ZenBook
상품명	3 15ACH R5 3050 (SSD 256GB)	Slim5-15ITL Premiere Plus i5 (SSD 256GB)	UM425UAZ-KI028 (SSD 512GB)
최저가격	930,000원	829,000원	1,130,860원
등록년월	2021.08	2021.01	2021.07
화면 크기	39.6cm(15.6인치)	39.6cm(15.6인치)	35.56cm(14인치)
CPU 제조사	AMD	인텔	AMD
CPU 종류	라이젠5-4세대	코어i5-11세대	라이젠7-4세대
CPU 코드명	세잔	타이거레이크	루시엔
CPU 넘버	5600H (3.3GHz)	i5-11350G7 (2.4GHz)	5700U (1.8GHz)
코어 수	핵사코어	퀵드코어	옥타코어
운영체제(OS)	미포함(프리도스)	미포함(프리도스)	미포함(프리도스)
화면 비율	와이드 16:9	와이드 16:9	와이드 16:9
해상도	1920x1080(FHD)	1920x1080(FHD)	1920x1080(FHD)
화면 원가	250nit	300nit	400nit



(1,1) 제조사/브랜드	(1,2) 레노버 게이밍	(1,3) 레노버 아이디어패드	(1,4) ASUS ZenBook
(2,1) 상품명	(2,2) 3 15ACH R5 3050 W10	(2,3) Slim5-15ITL Premiere i7	(2,4) UM425UAZ-KI028
(3,1) 최저가격	(3,2) 1,048,950원	(3,3) 998,950원	(3,4) 1,097,050원
(4,1) 등록년월	(4,2) 2021.07	(4,3) 2021.01	(4,4) 2021.07
(5,1) 화면 크기	(5,2) 39.6cm(15.6인치)	(5,3) 39.6cm(15.6인치)	(5,4) 35.56cm(14인치)
(6,1) CPU 제조사	(6,2) AMD	(6,3) 인텔	(6,4) AMD
(7,1) CPU 종류	(7,2) 라이젠5-4세대	(7,3) 코어i7-11세대	(7,4) 라이젠7-4세대
(8,1) CPU 코드명	(8,2) 세잔	(8,3) 타이거레이크	(8,4) 루시엔
(9,1) CPU 넘버	(9,2) 5600H (3.3GHz)	(9,3) i7-1165G7 (2.8GHz)	(9,4) 5700U (1.8GHz)
(10,1) 코어 수	(10,2) 핵사코어	(10,3) 퀵드코어	(10,4) 옥타코어
(11,1) 운영체제(OS)	(11,2) 윈도우10	(11,3) 미포함(프리도스)	(11,4) 미포함(프리도스)

한글 인코딩 방식 utf-8로 통일

<table> 태그로 테이블 재구성

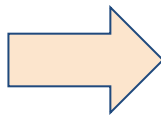
자연어 질의에서 의미없는 행 삭제

셀 좌표를 css로 삽입

## 테이블 데이터 전처리

학습이 어려운 레이아웃 변경

(0,0) 디스플레이	(0,1) • 화면크기 • 189cm	(0,2) • 해상도 • 3,840 x 2,160	
(1,0) 영상	(1,1) • 화질엔진 • Crystal Processor 4K	(1,2) • HDR (High Dynamic Range) • HDR	(1,3) • HDR 10+ • 지원



(0,0) 디스플레이	(0,1) 화면크기	(0,2) 189cm
	(1,0) 해상도	(1,1) 3,840 x 2,160
	(2,1) 화질엔진	(2,2) Crystal Processor 4K
	(3,0) HDR (High Dynamic Range)	(3,1) HDR
	(4,0) HDR 10+	(4,1) 지원

변경 전  
EM:24.1

변경 후  
EM:40.1



## 질의응답 데이터 수집 과정

번호	카테고리	질문유형	질문	답변	정답 위치
1	노트북	어휘 포함	ASUS ZenBook 13 - UM433DA의 메모리는 몇기 가예요?	8GB	(11,1)
2	무선 이어 폰	어휘 변형	airpod pro는 어떤 버전의 블루투스를 탑재했는가?	블루투스 v5.0	(4,3)
3	태블릿	다중 근거	탭4 8 플러스, 미디어패드 M6 프로 8.4, 미 패드4는 CPU의 코어 개수가 몇 개인가요?	옥타코어	(12,1)
4	TV	다중 근거	N651UHD VA, 쿠카 UC861UHD, KU75UA8100F 중에서 에너지효율이 1등급인 상품명은?	N651UHD VA&&KU75UA8100F	(1,1)&&(1,3)
5	스마트폰	대소 비교	Apple iPhone 8, Apple iPhone 7, Apple iPhone 6 중 가 장 무거운 제품명은?	Apple iPhone 8	(0,1)
...	...	...	...	...	...
3065	스마트밴드	순서	... 중 가장 나중에 등록된 제품명은?	미밴드6	(1,1)

## 어휘 포함, 어휘 변형

저장 용량	<ul style="list-style-type: none"> <li>• 64GB</li> <li>• 128GB</li> </ul>
방수 및 방진	<ul style="list-style-type: none"> <li>• IEC 규격 60529하의 IP68 등급 획득(최대 수심 2m, 최대 30분)</li> </ul>
칩	<ul style="list-style-type: none"> <li>• A13 Bionic 칩</li> <li>• 3세대 Neural Engine</li> </ul>

질문유형	질문
어휘 포함	iPhone 11가 지원하는 방수 및 방진등급은 무엇인가요?
어휘 변형	iPhone 11에는 먼지가 들어오는 걸 막는 기능이 있나요?

- 어휘 변형 질문에는 전자기기에서 사용하는 어휘를 변형한 질문 포함(예: 화질, 해상도)

## 다중근거, 대소 비교, 순서

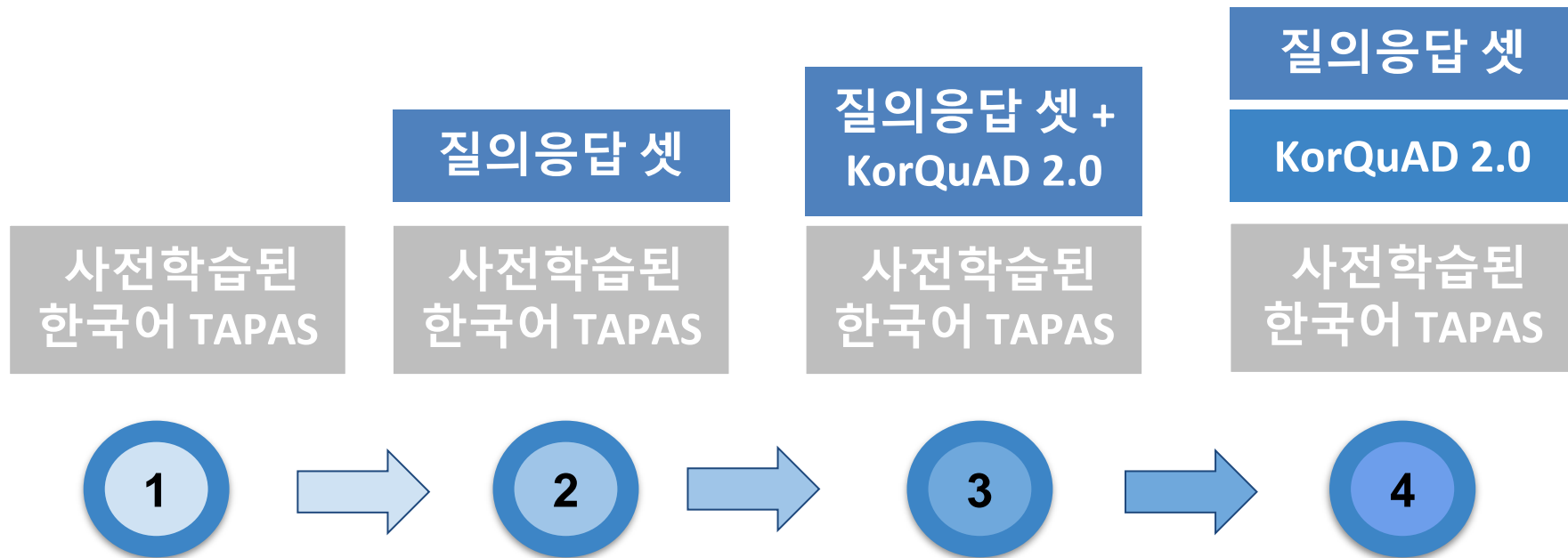
상품명	Inspire 2	M3 PLUS	Charge 4 SE
사용시간	평균사용:10일 이하	평균사용:7일 이하	평균사용:7일 이하
연결	블루투스 NFC	블루투스	GPS NFC
등록년월	2020.09	2021.02	2020.04



다중 근거, 대소 비교, 순서 질문은 다중 테이블에서 수집

질문유형	질문	답변
다중 근거	Inspire 2, M3 PLUS, Charge 4 SE 중에서 연결 기능으로 GPS를 지원하는 제품의 상품명은 무엇인가?	Charge 4 SE
다중 근거	... 중 평균사용 사용시간이 7일이하인 상품명은 무엇인가?	M3 PLUS && Charge 4 SE
대소 비교	... 중 평균사용 이용시간이 가장 긴 상품명은 무엇인가?	Inspire 2
순서	... 중 등록년월이 가장 최신인 상품명은 무엇인가?	M3 PLUS

## 추가 학습 (fine tuning) 진행



## 연구 결과 분석 및 의의

- 연구 결과 나타난 데이터에 대한 분석과 연구의의에 대해 알아봅니다.

## 단일, 다중 테이블 연구 결과 분석

### 언어 모델 성능 지표

**F1 score**

정밀도와 재현율을 종합적으로 고려한 수치

셀을 정확히 찾은 경우 증가한다.

**EM score**

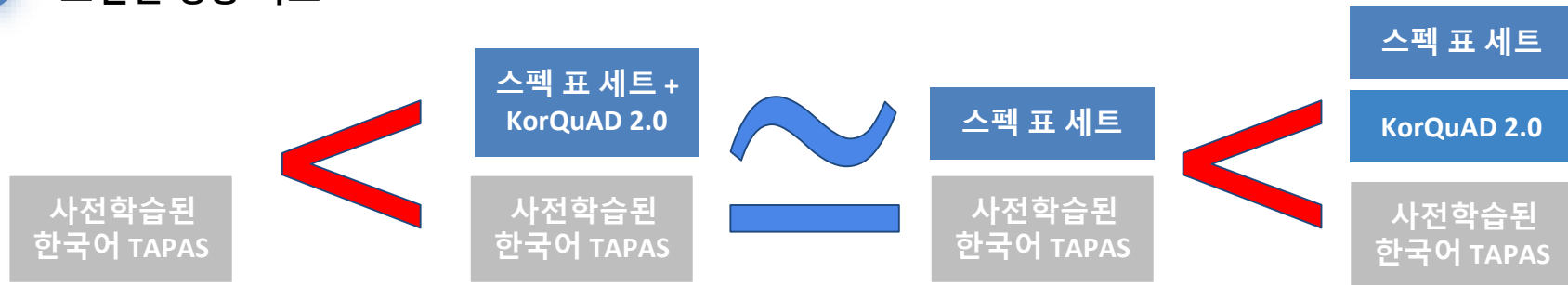
모델이 정확하게 정답을 맞춘 비율

### 모델별 성능 비교

성능 지표	단일 테이블		다중 테이블	
	F1	EM	F1	EM
한국어에 맞게 바꾼 TAPAS	56.9	49	40.0	18.6
KorQuAD 2.0와 스펙 표 질의응답 데이터를 함께 파인 튜닝	58.2	49.2	39.0	13.9
스펙 표 질의응답 데이터로 파인 튜닝	81.5	73.6	66.1	48.1
KorQuAD 2.0로 파인 튜닝 후 스펙 표 질의응답 데이터로 파인 튜닝	81.9	75.2	77.4	64.3

## 파인튜닝 방식에 따른 모델별 연구 결과 분석

### 모델별 성능 비교



01

파인튜닝과정에서 이번 연구에서 제작한 스펙표 세트를 사용하였을때 한국어 제품 스펙 표에 대한 질의응답에서 25퍼센트 정도의 유의미한 정답률 상승을 보였다.

02

스펙 표 데이터셋과 KorQuAD 2.0을 함께 파인튜닝한 결과가 따로 파인튜닝한 것보다 낮은 성능을 보였다. 이는 서로 표의 형태가 크게 달라서 같이 학습을 시키니 스펙 표가 오히려 노이즈로 작용하여 따로 학습한 모델에 비해 학습이 덜 된것으로 보인다.

## 테이블 비율별 모델 연구 결과 분석

- 테이블 비율별로 파인튜닝을 추가진행(학습 셋 70%, 테스트 셋 30%)

	A		B	
성능 지표	F1	EM	F1	EM
KorQuAD 2.0와 스펙 표 질의응답 데이터를 함께 파인 튜닝	74.8	63.3	64.0	50.6
스펙 표 질의응답 데이터로 파인 튜닝	74.5	62.7	64.0	50.5
KorQuAD 2.0로 파인 튜닝 후 스펙 표 질의응답 데이터로 파인 튜닝	81.5	72.5	74.5	64.3

A

수집한 표 70%

표에 대해 수집한  
질의응답  
100%(1203개)

B

수집한 표 100%

표에 대해 수집한  
질의응답  
70%(1276개)

A 모델이 더 높은 성능을 보였다. 이는 테이블에 더 자세히 질문을 하는 게 학습이 잘됐음을 보여준다.

추후 연구에서 더 많은 데이터셋을 수집하면 더 정확히 알 수 있을 것으로 보인다.



## 질문 유형별 연구 결과 분석

※ A모델, KorQuAD 2.0로 파인 튜닝 후  
스펙 표 질의응답 데이터로 파인 튜닝

성능 지표	F1	EM	개수
전체	81.5	72.5	1182
어휘 포함	90.2	85.0	294
어휘 변형	84.2	78.3	249
다중 근거	83.3	77.3	362
대소 비교	69.3	49.4	259
순서	66.8	50.0	18

가장 결과가 잘 나온 모델인 **A 모델의 파인 튜닝을 두 번** 한 경우로 질문 유형별로 테스트를 진행했다

EM수치 기준으로 **어휘 포함**과 **다중 근거**가 **높게** 나왔다. 단일 제품 표와 다중 제품 표에서 셀의 이름을 직접 질의하여 비교적 쉽게 정답을 도출한 것으로 보인다.

EM수치 기준으로 **대소 비교**는 49.4, **순서**는 50.0로 **낮게** 나왔다. 대소 비교는 다중 근거 대비 EM수치가 떨어짐을 확인하였고, 순서 또한 데이터셋 수가 적기는 하지만 다중 근거 대비 EM수치가 나오지 않음을 알 수 있다.

# 04 | 연구 결과 분석 및 의의

## 다중테이블 유형의 파인튜닝 전후 결과 비교

성능지표	F1	EM	개수	F1-EM
전체	47.9	34.4	1182개	13.5
다중 근거	32.3	17.6	362개	14.6
대소 비교	42.9	16.2	259개	26.7
순서	25.3	5.5	18개	19.8

※ 사전학습 모델로 테스트 진행

정답 셀 선택



정답 추출 실패



F1 - EM

질문	실제 정답	추출한 정답
가로의 길이가 가장 긴 상품명은?	갤럭시탭 프로S 12.0 코어M3 6세대	코어M3
해상도 크기가 가장 큰 상품명은?	갤럭시탭 프로S 12.0 코어M3 6세대	갤럭시탭 프로S
코어클럭의 최대값이 가장 작은 상품명은?	갤럭시탭 프로S 12.0 코어M3 6세대	갤럭시탭 프로S 12.0 코어M3
RAM이 가장 작은 상품 명은?	갤럭시탭 프로S 12.0 코어M3 6세대	12.0

제품명이 일반적으로 인명이나 사물의 이름과는 달리 여러개의 단어로 구성되어 정답 추출이 어렵다.

## 다중테이블 유형의 파인튜닝 전후 결과 비교

성능지표	F1	EM	개수	F1-EM
전체	47.9	34.4	1182개	13.5
다중 근거	32.3	17.6	362개	14.6
대소 비교	42.9	16.2	259개	26.7
순서	25.3	5.5	18개	19.8

성능지표	F1	EM	개수	F1-EM
전체	81.5	72.5	1182개	9.0
다중 근거	83.3	77.3	362개	6.0
대소 비교	69.3	49.4	259개	19.9
순서	66.8	50.0	18개	21.1

제품명의 경계값을 잘 찾지 못  
해 F1과 EM의 차이가 크다

파인튜닝 결과 성능 향상

# 04 | 연구 결과 분석 및 의의

## 대소비교, 순서 연구 결과 분석

성능 지표	F1	EM	개수
전체	81.5	72.5	1182
대소 비교	69.3	49.4	259
순서	66.8	50.0	18

제품명	Samsung Galaxy M12	iPhone SE (2020)	LG X6 2019
크기(WxHxD,mm)	75.9 x 164 x 9.7	67.3 x 138.4 x 7.3	77 x 161.3 x 8.7
화면 해상도	720 x 1600 픽셀	750 x 1334 픽셀	720 x 1520 픽셀
출고가	₩198,000	₩605,000	₩349,800

대소 비교와 순서를 물어보는 질문들의 성능이 전체 성능에 비해 25퍼센트 정도 떨어진다.

다양한 단위 기호가 포함되어 있어 셀마다 대소비교를 정확히 매기기 어렵다.

## 연구 의의

- 연구가 덜 되어있는 분야인 한국어 스펙 표에 특화된 자연어 질의응답 데이터셋을 수집하고, TAPAS 모델이 스펙 표에 대해 정답을 잘 추출하지 못하는 케이스들을 찾았다.

- 스펙 표에 특성상, 기존에 비해 생소한 질의응답 유형을 만들고 파인튜닝을 통해 모델 성능이 개선되는것을 검증했다.

- 모델의 성능 향상을 위해 다양한 레이아웃을 가진 스펙 표를 모델이 잘 학습하도록 변경하는 과정을 수행했다.

- 스펙 표 질의응답 데이터셋을 사용해 모델을 파인 튜닝함에 있어서, 여러 모델을 만들어 비교해서 가장 학습을 잘하는 모델을 얻는 방법을 탐색했다.

Q&A