



과기정통부 혁신성장동력프로젝트

엑소브레인 한국어 언어이해 및 질의응답 기술 소개

(휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발)

2020. 11. 30.

임준호

언어지능연구실 / 한국전자통신연구원

발표 순서

- I 엑소브레인 과제 개요
- II 심층 언어분석 기술
- III 딥러닝 언어모델 기술
- IV 심층 질의응답 기술



언어를 이해하고, 지식을 학습하여, 지식을 서비스하는 AI 기술
- 인간과 기계의 Q&A를 통한 지식서비스 제공 -

<2012년>
국가 차원
인공지능
장기
프로젝트
필요성 제기

2013.05

1단계 (핵심 기술 개발)

목표
및
결과물

IBM 왓슨과 기술격차 7년을
4년간 개발하여 단축
(단답형 질의응답)

기술검증
장학퀴즈 <대결! 엑소브레인>
압승('16.11.18)



엑소브레인 인공지능
산/학/연 생태계 구축

엑소브레인 생태계



표준화, SW보급, DB/말뭉치

2017.03

2단계 (응용 기술 개발)

전문분야 질의응답
시스템
(서술형 질의응답)

전문지식 QA 시스템 상용화
(일반, 법률, 금융, 공공분야)

엑소브레인 오픈API



전문지식 QA



법률/특허 전문지식 제공

지능형 상담 QA



헬프데스크 상담 QA

2020.01

3단계 (글로벌 기술 개발)

전문가 의사결정
지원 시스템
(설명형 질의응답)

의사결정 지원 시스템 상용화

전문분야 컨설팅 QA



전문가 의사결정 지원

다국어 지식 QA



다국어 지식서비스 제공

지능형 로봇 QA



스마트머신 지능적 업무 수행

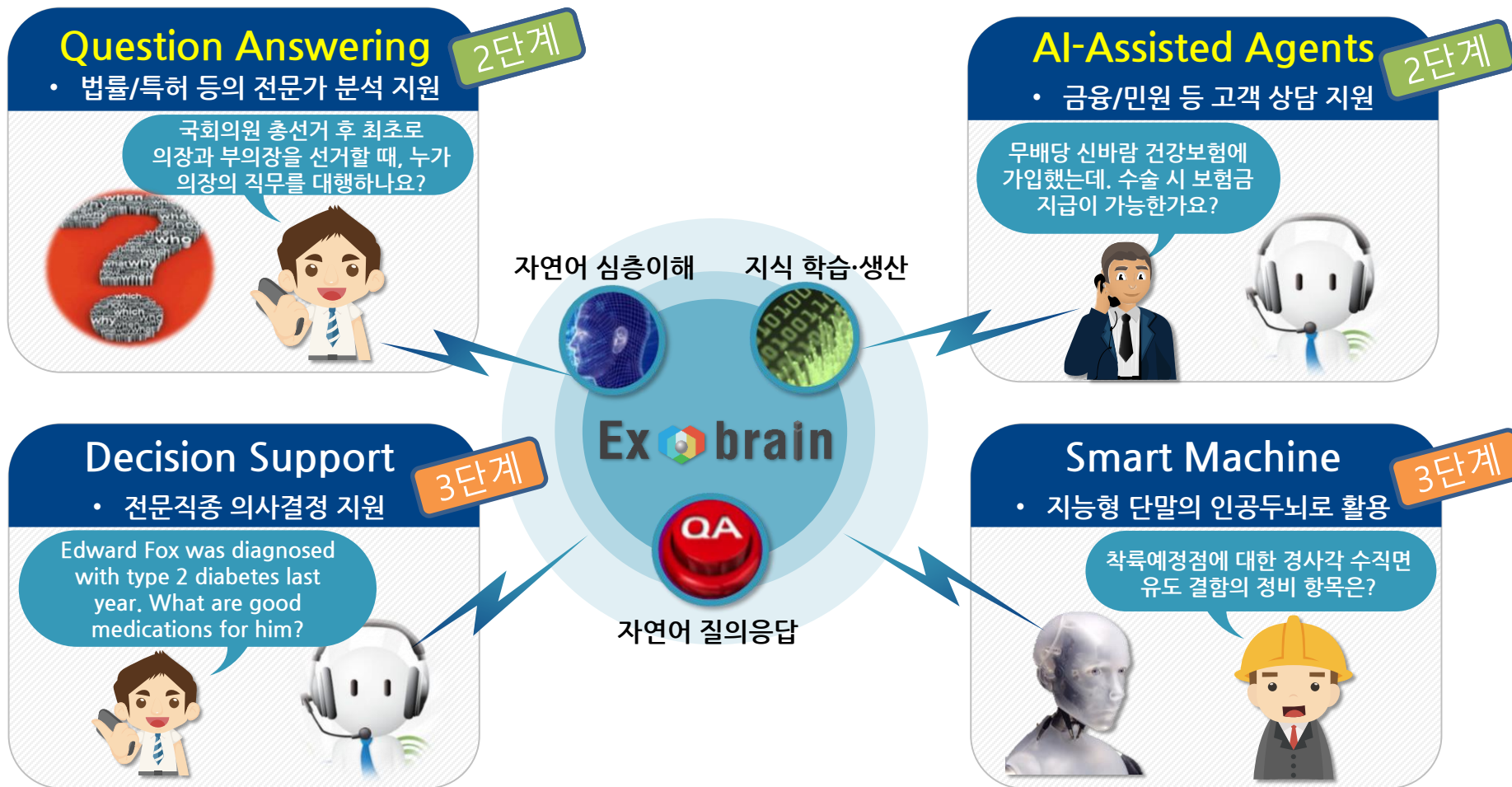
2023.02

기술검증
및
사업화



전문가(예: 변호사, 변리사)와 질의응답이 가능한 인공지능 SW 개발

* 전문분야에서 질의응답(Question Answering)을 통한 지식 제공 서비스





자연어로 기술된 **문제의 의미를 이해하고, 정답을 추론하여 생성함**

- 문제: 국회에서 긴급한 경우 본회의 개의 방법은?
- 정답: 회의의 일시만을 의원에게 통지하고 개의

정보검색: 단어 → 문서

Google search results for '국회에서 긴급한 경우 본회의 개의 방법은?'. The top result is from the National Legislation Information Center (www.law.go.kr) with the title '법령 > 본문 - 국회법 - 국가법령정보센터'. The snippet reads: 'Jul 17, 2018 - ② 국회는 휴회 중이라도 대통령의 요구가 있을 때, 의장이 긴급한 필요가 ... ① 의법에서 정한 사항 외에 국회의사처에 관한 사항은 따로 법률로 정한다. ... 아니하는 경우에는 그 이후에 최초로 개의하는 본회의에 상정하여 표결한다.' A red arrow points from this snippet to the '국회법 - 국가법령정보센터' link below.

질의응답: 질문 → 정답

exobrain Q&A system interface. The question is: '질문: 국회에서 긴급한 경우 본회의 개의 방법은?'. The answer is categorized under '① 회의의 일시만을 의원에게 통지하고 개의' and includes the following text: '국회법 | 제6장 회의 / 제1절 개의·산회와 의사일정 | 제76조 의사일정의 작성' followed by numbered points ① through ⑤ explaining the procedure for urgent sessions. A second category '② 재개' is also visible below.



워드마크를 삽입하려면?

exobrain
법률 질의응답 시스템

언어를 이해하고, 지식을 학습하여, 법률지식을 제공하는 질의응답 시스템
(본 시스템은 Chrome 브라우저 사용 권장입니다.)

◆ 검색할 질문

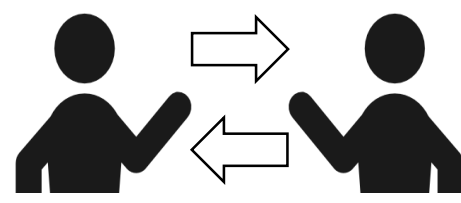
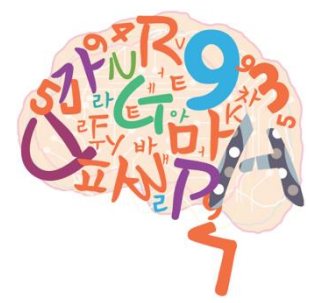
1. 검색할 질문이 무엇?
2. 질문을 작성할 때 주의할 사항은?
3. 답변과 영수?
4. 영수증 발급을 위해 목적은 무엇인가?
5. 영수증 발급이 가능한가?

◆ 답변할 질문

1. 해당 질문을 읽고 답을 작성할 것인가?
2. 답변할 때 필요한 지식은 무엇인가?
3. 답변할 때 필요한 지식은 어디에서 얻을 것인가?
4. 답변할 때 필요한 지식은 어디에서 얻을 것인가?
5. 답변할 때 필요한 지식은 어디에서 얻을 것인가?

◆ 사용할 질문

1. 사용 가능한 질문은 어떤 것인가? (필수 질문)
2. 사용 가능한 질문은 어떤 것인가? (필수 질문)
3. 사용 가능한 질문은 어떤 것인가? (필수 질문)
4. 사용 가능한 질문은 어떤 것인가? (필수 질문)
5. 사용 가능한 질문은 어떤 것인가? (필수 질문)



(1) 기술완성도 뒷받침 시,
산업적 활용도 높음

(2) 사람과 같이
질문을 이해하고 정답을
추론하는 과정에 필요한
축적되는 기술이 많음

(3) 질문&답변 형태는
**(인공)지능 수준 평가 및
실세계 적용의
가장 직관적 형태**

사내 기술문서, 규정집 등 문서를
검색하거나, 필요한 지식을 찾는
모든 활동에 적용 가능한 파급력
이 큰 기술

- 현재수준: 파일 이름 기반 검색
- most active player: google

컴퓨터가 언어에 포함된
의미를 이해하고,

여러 문장 사이의 관계를
이해하는 기술 확보

단일 근거 이해 수준,

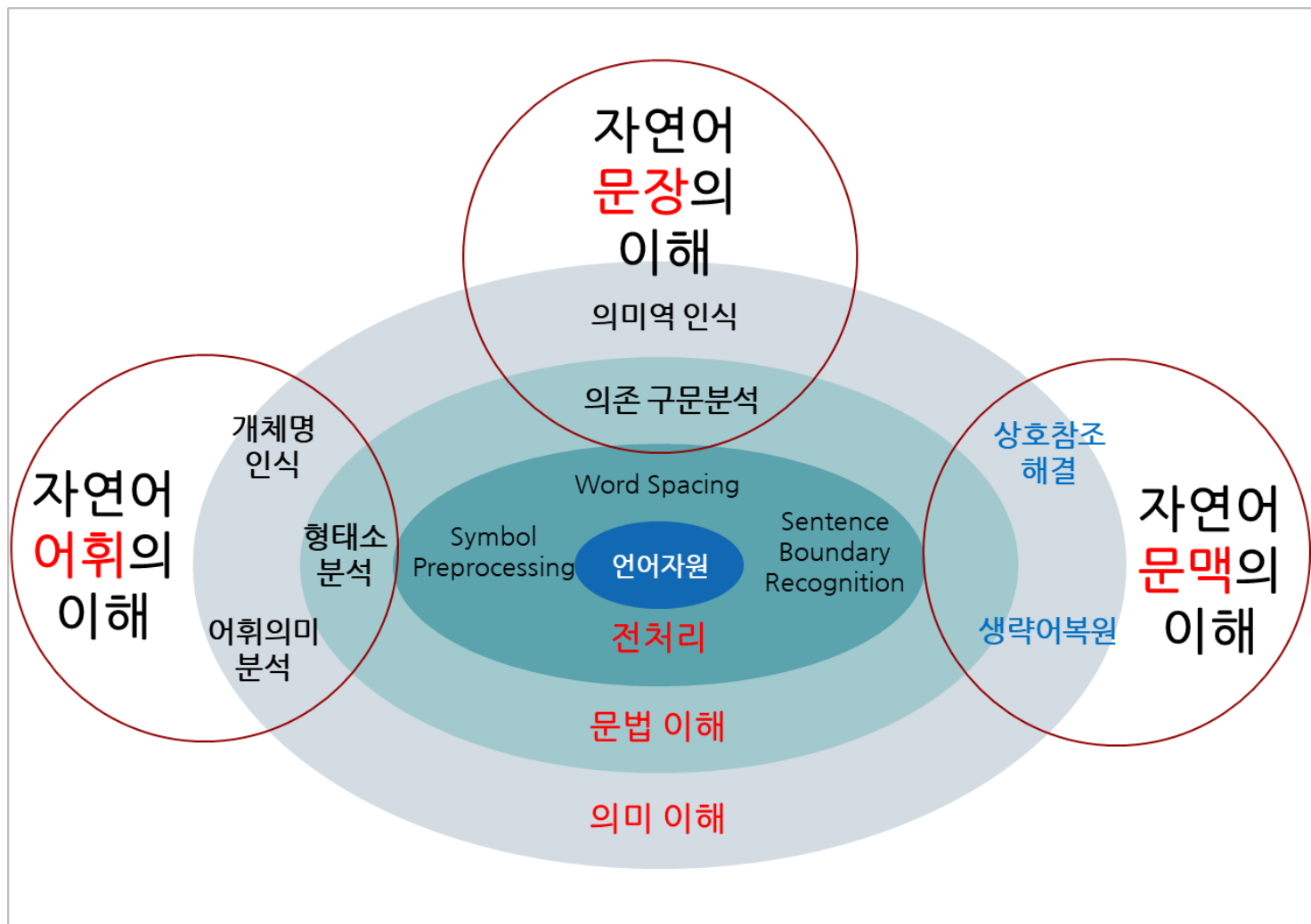
복수 근거간 관계 이해 수준,
상식 이해 수준,
멀티모달 이해 수준

발표 순서

- I 엑소브레인 과제 개요
- II 심층 언어분석 기술
- III 딥러닝 언어모델 기술
- IV 심층 질의응답 기술



- 텍스트에 기술된 언어의 뜻과 의미를 분석하는 기술
 - 사람이 언어를 이해하는 방식 모사



(1) 심층 언어분석 (어휘/문장/문맥 이해)



소피스트란 그리스어로 지혜로운 자 또는 지혜를 만들어내는 사람이라는 뜻으로, BC 5~4세기의 그리스의 철학자들을 말한다. 이들은 아테네 사람들을 대상으로 하였고, 수사학과 웅변술을 가르쳤다.

형태소 분석

소피스트/NNG+란/JX 그리스/NNP+어/XSN+로/JKB 지혜롭/VA+ㄴ/ETM 자/NNB 또는/MAG 지혜/NNG+를/JKO 만들/VV+어/EC+내/VX+는/ETM 사람/NNG+이/VCP+라는/ETM 뜻/NNG+으로/JKB+,/SP BC/SL 5/SN+~/SO+4/SN+세기/NNP+의/JKG 그리스/NNP+의/JKG 철학/NNG+자/XSN+들/XSN+을/JKO 말/NNG+하/XSV+s다/EF+./SF 이/NP+들/XSN+은/JX 아테네/NNP 사람/NNG+들/XSN+을/JKO 대상/NGG+으로/JKB 하/VV+았/EP+고//EC+./SP 수사/NNG+학/XSN+과/JC 웅변/NNG+술/XSN+을/JKO 가르치/VV+았/EP+다/EF+./SF

개체명 인식

<CV_OCCUPATION:소피스트/NNG>+란/JX <CV_LANGUAGE:그리스/NNP+어/XSN>+로/JKB 지혜롭/VA+ㄴ/ETM 자/NNB 또는/MAG 지혜/NNG+를/JKO 만들/VV+어/EC+내/VX+는/ETM 사람/NNG+이/VCP+라는/ETM 뜻/NNG+으로/JKB+,/SP <DT_DURATION:BC/SL 5/SN+~/SO+4/SN+세기/NNP>+의/JKG <LCP_COUNTRY:그리스/NNP>+의/JKG 철학/NNG+자/XSN+들/XSN+을/JKO 말/NNG+하/XSV+s다/EF+./SF 이/NP+들/XSN+은/JX <LCP_CAPITALCITY:아테네/NNP> 사람/NNG+들/XSN+을/JKO 대상/NGG+으로/JKB 하/VV+았/EP+고//EC+./SP <FD_ART:수사/NNG+학/XSN>+과/JC <FD_ART:웅변/NNG+술/XSN>+을/JKO 가르치/VV+았/EP+다/EF+./SF

어휘의미분석

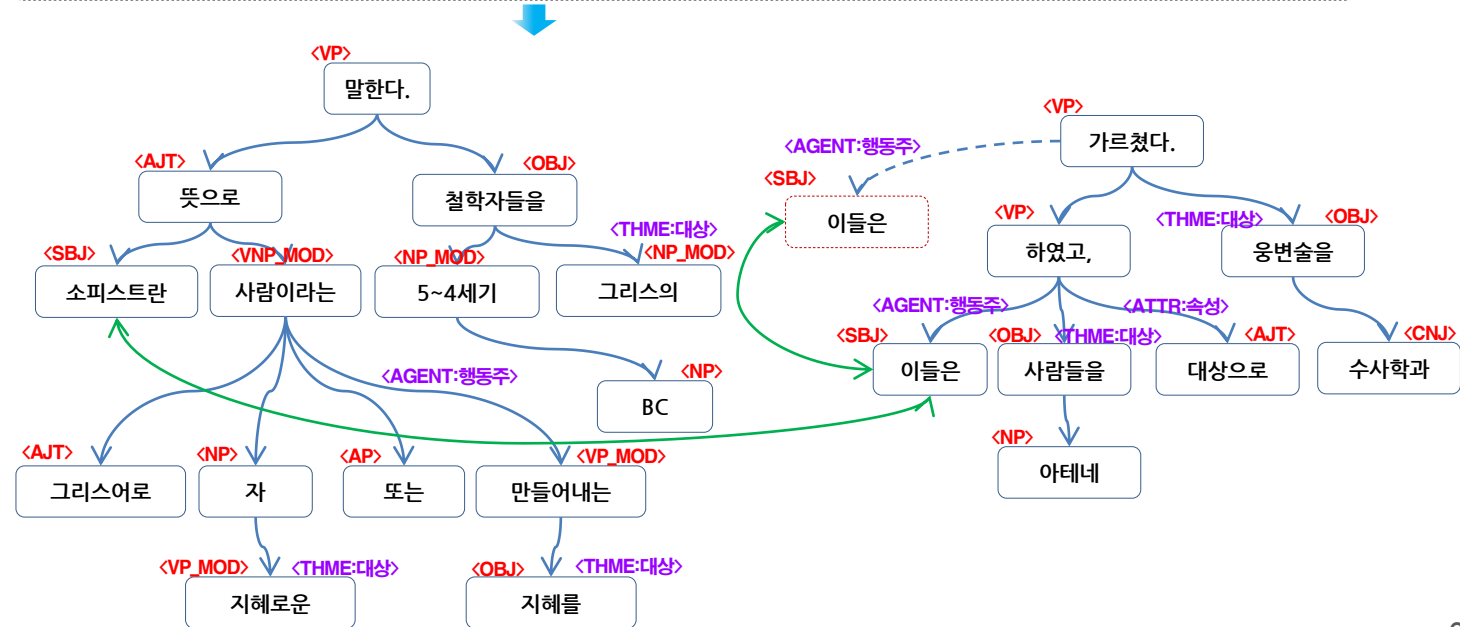
소피스트란 그리스어로 지혜로운 자_18_0000/NNB 또는 지혜_02_0001/NNG+를 만들_00_0101/VV+어내는 사람_00_0001/NNG+이라는 뜻_00_0002/NNG+으로, BC 5~4+세기_03_0002/NNG+의 그리스_02_0000/NNP+의 철학자들을 말하_00_0101/VV+는다.

의존구문분석

의미역 인식

상호참조해결

무형대용어 생략복원



(1) 심층 언어분석 (어휘/문장/문맥 이해)



(1) 오픈API를 이용한 기술공개



1,371
개 기관

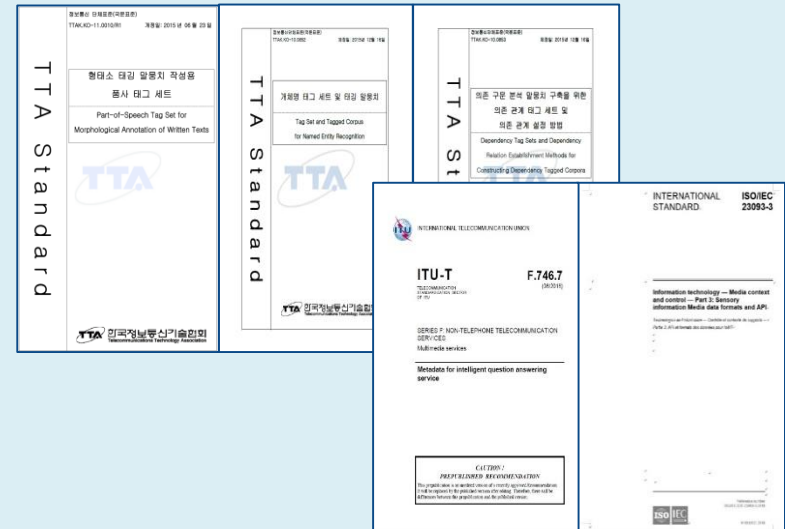
2천 4백
만건
(일 평균
2만8천건)

(2) 언어분석 기술이전 19건

- 한글과컴퓨터, 마인즈랩, 금융분야 M사, H사, 등

(3) 국내외 표준화 추진

- TTA 국내표준 7건 승인
- ISO 및 ITU-T 국제 표준 4건 승인



(4) 국립국어원 국어 빅데이터 구축 사업에 태깅가이드, 언어자원 제공 (200억원 규모)

- 한국어 언어분석 7종 태깅가이드 제공 (ETRI 국내 표준안 4건 포함)
- 한국어 의미역 프레임셋 제공

(1) 심층 언어분석 (어휘/문장/문맥 이해)

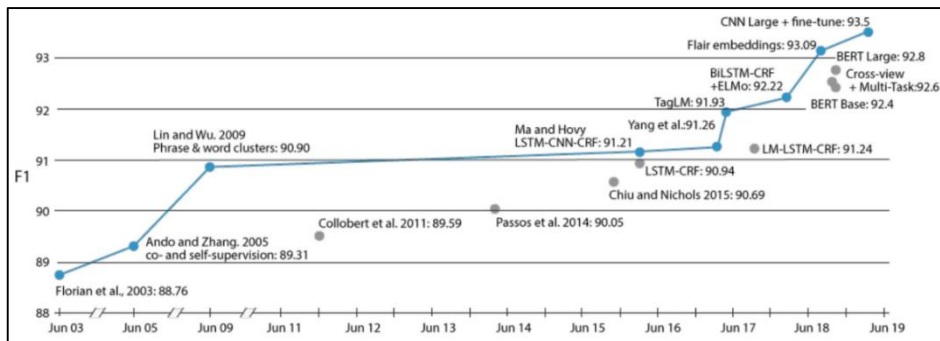
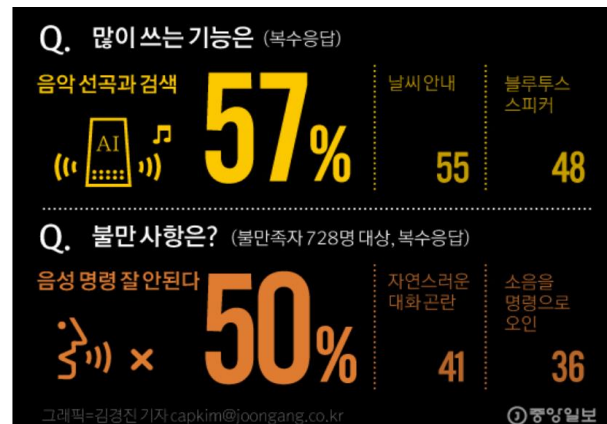
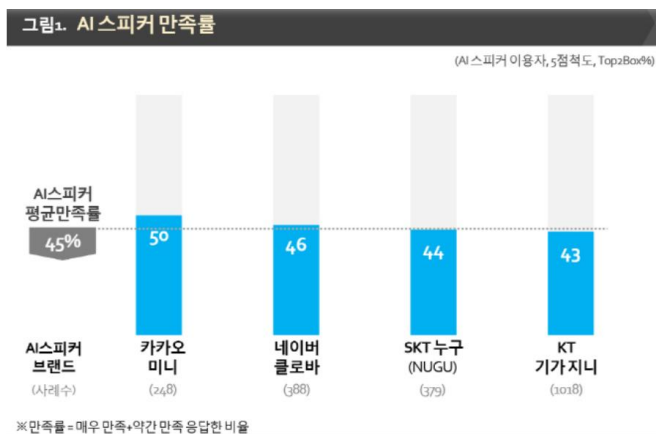


• 기술동향

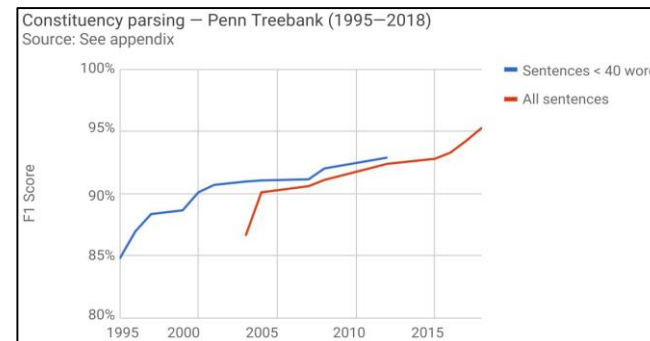
- 문어(글) 위주의 자연어 이해에서 구어(말)로 확장 필요

- 영어 트윗 대상 형태소 분석 최고 성능: 91.18% (Fudan University)
- 영어 웹/방송 텍스트 대상 개체명 인식 최고 성능: 84.97% (7개 클래스 / CMU)

- 참고: AI 스피커 만족도 조사(출처: 중앙일보, 19.1)



〈개체명 인식: '09년 90.9%, '19년 93.5%〉



〈의존 구문 분석: '10년 92%, '19년 96%〉

(1) 심층 언어분석 (어휘/문장/문맥 이해)

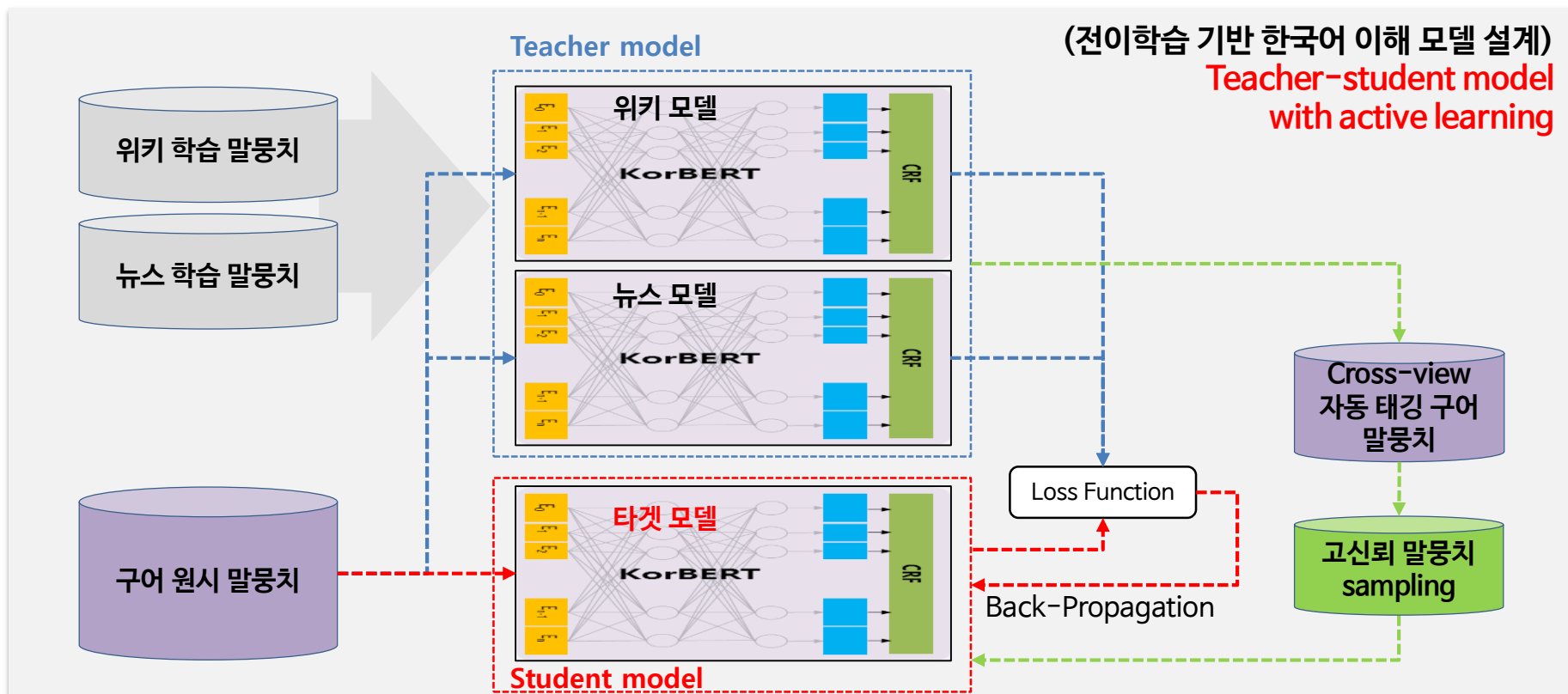


연구 목표

- 소량의 학습데이터만으로도 **도메인 확장**이 가능한 **전이학습 기반 한국어 이해 기술** 개발
- 문어/구어 형태소분석 및 개체명인식 기술 개발
- **OpenAPI 서비스**를 통한 외부 공개

연구 현황 및 계획

- 전이학습 기반 한국어 이해 모델 설계
- 형태소분석 문어/구어 성능 현황
 - 문어: 96.80%, 구어: 92.95% (어절정확도)
- 개체명인식 문어/구어 성능 현황
 - 문어: 89.40%, 구어: 88.30% (146개 세분류 성능)

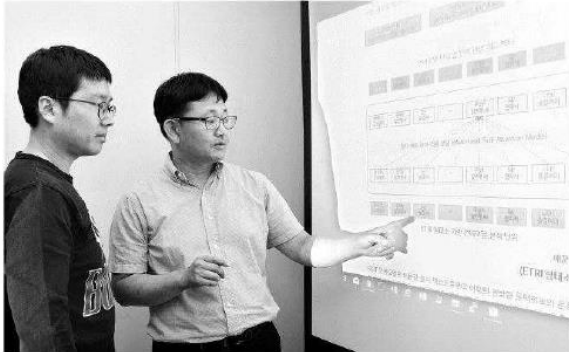


(1) 심층 언어분석 (어휘/문장/문맥 이해)



디지털타임스

2020년 08월 07일 금요일
010면 신업과학 11.5 x 22.8 cm



ETRI 연구진들이 6일 언어 인식 인공지능 '엑소브레인'을 기반으로 사람의 말까지 이해할 수 있는 '구어체 언어분석 기술'을 개발, 일반에 공개했다. ETRI 제공

구어체까지 이해하는 AI 나왔다

ETRI 일상언어 분석기술 개발 음성인식 AI 서비스 진화 전망

국내 연구진이 AI(인공지능) 기술을 적용해 사람의 일상적인 대화까지 정확하게 분석해 내는 기술을 개발했다. 글을 이해하는 AI를 넘어 사람의 말까지 이해하는 다양한 AI 서비스 개발로 이어질 전망이다.

한국전자통신연구원(ETRI)은 자체 개발한 언어 AI '엑소브레인'의 '구어체 언어분석 API(응용 프로그래밍 인터페이스)'를 공개했다고 6일 밝혔다. 이 기술은 위키백과와 법령 등 문어체를 주로 이해하는 기존 언어분석 기술을 확장시킨 것으로, 사람의 대화 분석 오류를 최대 41%까지 줄였다.

구어체 언어분석 기술은 한국어 의미의 최소 단위를 분석하는 '형태소분석 기술'과 문장 내 고유 대상과 의미를 인식하는 '개체명 인식기술'을 적용했다. 구어체 언어분석은 난이도와 학습 데이터 부족으로 구현하기 어렵

다. 가령, '경상도인데'를 '경상도네'라고 구어체로 표현하는 경우, 기존 형태소분석 기술은 '경상도+인데'라는 축약표현을 인식하지 못해 '경상도+네'로 잘못 분석한다.

연구팀은 전이학습과 데이터 증강 기법을 활용해 이 같은 문제를 극복했다. 전이학습과 데이터 증강 기법은 다른 분야의 학습 모델과 소량의 학습 데이터를 재사용해 학습하는 방식으로, 학습 데이터가 부족한 환경에서 딥러닝 기술 한계를 극복하는 데 활용된다. 두 기술을 적용한 결과, 구어체 언어분석 API는 기존 문어체 언어분석 API와 비교해 형태소 분석과 개체명 인식 성능이 각각 5.0%, 7.6% 개선된 것으로 나타났다. 분석 오류도 각각 41.7%, 39.4% 감소했다.

임준호 ETRI 언어지능연구실 박사는 "기존 문어체 기술을 고도화해 사람의 대화까지 정확하게 이해할 수 있는 구어체 언어분석 기술 개발을 통해 AI 비서, 챗봇 등 AI 서비스 개발에 기여할 것"이라고 말했다. 이준기기자

참고1 엑소브레인 언어분석 API 성능

		평가셋	오픈소스 ¹⁾ (mecab)	기존 OpenAPI	'20.08월 공개 OpenAPI	오류 감소율
문어체	형태소분석	4,800문장	87.36%	95.09%	96.80% (+1.7%)	34.83%
	개체명인식	2,320문장	-	87.80%	89.40% (+1.6%)	13.11%
구어체	형태소분석	4,800문장	82.34%	87.90%	92.95% (+5.0%)	41.74%
	개체명인식	2,323문장	-	80.70%	88.30% (+7.6%)	39.38%

(성능 평가 기준)

- * 형태소분석: 정보통신 단체표준(국문표준) TTAK.KO-11.0010/R1 "형태소 태깅 말뭉치 작성용 품사 태그 세트 가이드라인" 기준 어절 단위 정확률 평가
- * 개체명인식: 정보통신 단체표준(국문표준) TTAK.KO-10.0852 "개체명 태그 세트 및 태깅 말뭉치" 기준 146개 세분류 대상 F1 평가

¹⁾ 오픈소스로 형태소분석은 mecab 라이브러리와 비교 및 개체명인식은 비교 대상 오픈소스 미확인

발표 순서

- I 엑소브레인 과제 개요
- II 심층 언어분석 기술
- III **딥러닝 언어모델 기술**
- IV 심층 질의응답 기술



딥러닝 기반 언어 문맥 학습(contextual representation) 기술

- * 한국어 = 내용어(명사, 동사 등)와 기능어(조사, 어미 등)이 결합하여 어절을 구성하는 교착어
- * 한국어에 최적화된 언어 모델 개발 (구글 대비 4.5% 우수)

	구글 배포 모델 언어 단위	ETRI KorBERT 언어 단위
접근 방법	영어*의 언어 단위 구성 방법 적용 * 내용어와 기능어가 공백으로 구분된 언어 (굴절어)	한국어* 의미의 최소 단위인 형태소 기반 언어단위 * 내용어와 기능어가 어절로 결합된 언어 (교착어)
언어 단위 구축 방법	말뭉치에서 통계적으로 추출한 음절 단위 n-gram 적용 방법 (BPE)	형태소 분석 이후, 형태소 단위에 대해서 BPE 적용 방법
언어 단위 적용 결과	한국 ##어 단 ##어는 형태 ##소로 구 ##성된 ##다.	한국어/NNP_ 단어/NNG_ 는/JX_ 형태소/NNG_ 로/JKB_ 구성/NNG_ 되/XSV_ ㄴ다/EF_ /SF_
동일 학습데이터 평가결과	Google wordpiece 모델 - 기계독해: 90.68% - 단락순위화: 66.3%	KorBERT 형태소 모델 - 기계독해: 95.02% (+4.3%) - 단락순위화: 73.7% (+7.4%)

※ BERT 언어모델에서 언어 단위의 중요성

1. BERT 언어모델 = 입력 언어 단위 사이의 관계 학습 모델

- (구글방식) 주식 문서의 “주가”와 “백설공주가”의 “주가” 사이의 모호성(ambiguity) 증가

2. 입력 언어 단위 = 출력 언어 단위 → 언어 단위 별 응용 태스크 활용

- (구글방식) “백설공주가” 에서 “주가”를 한 단위로 인식 → “주가” 단위로 개체명인식 활용

(2) 딥러닝 언어모델 (KorBERT)



- (1) 언어모델 학습 고도화를 위한 **자가학습 및 가중치 학습 방법 고도화**
- (2) 전이학습 기반 **도메인 특화 법률 · 특허 언어모델 개발 및 성능 개선**

[일반분야 언어모델 구축]

대용량 한국어 텍스트

- 위키백과 및 신문기사 23.5 GB (약 15년 분량)
- 47억개 형태소 수집

※ 자가학습 태스크 및 가중치 학습 방법 비교

KorBERT 모델	구글 배포 모델
Adaptive difficulty level (15% → 20%)	Fixed difficulty level
Whole word masking	Single token masking
Dynamic masking	Static masking
Robust Optimization (8,000 example base)	Regular optimization (256 example base)

- ※ 학습 대상 예문 집합 중, “눈/명사”가 모두 eye 의미 경우,
- Snow 의미에 대한 올바른 가중치 학습 실패 가능성 증가
- 256개 예문 기반 학습 시, 데이터 편향 위험성 증가
- 8,000개 예문 기반 학습 시, 올바른 가중치 학습 가능

학습 대상 문서 선별

형태소 기반 의미화
(47억개 형태소)

학습 시 동적 문제 생성
(신규 문제, 난이도 상향)

언어 모델 예측

대량 예제 기반
올바른 가중치 학습
(학습 안정성 향상)

KorBERT 언어모델

[법률/특허 도메인 언어모델 구축]

일반분야 KorBERT
+ 법률분야 텍스트
(186MB)

특허분야
원시 텍스트
(XML 53GB)

학습 대상 문서 선별

형태소 기반 의미화

법률/특허 분야
문제 생성

언어 모델 예측

대량 예제 기반
올바른 가중치 학습

KorBERT-Legal
언어모델

KorBERT-Patent
언어모델

법률 기계독해 적용
→ 2.01% 개선

특허 분류 적용
→ 6.25% 개선

(2) 딥러닝 언어모델 (KorBERT)



- (1) 외부 공개 및 산업계/학계 주요 딥러닝 언어모델로 자리매김 (논문 citation)
- (2) 구글 BERT 모델 대비 평균 4.5% 우수 (5개 태스크 대상 비교 평가)
- (3) '19.10 한국어 기계독해 챌린지 1위 (LG CNS 주관)
- (4) 딥러닝 언어모델 기술이전 4건 (2.4억, '20.05~)

KorQuAD 1.0

The Korean Question Answering Dataset

Leaderboard

KorQuAD 1.0의 Test set으로 평가한 Exact Match(EM) 및 F1 score입니다.

Rank	Reg. Date	Model	EM	F1
-	2018.10.17	Human Performance	80.17	91.20
1	2019.10.25	KorBERT-Large v1.0 ETRI ExoBrain Team	87.76	95.02
2	2019.06.26	LaRva-Kor-Large+ + CLaF (single) Clova AI LaRva Team	86.84	94.75
3	2019.06.04	BERT-CLKT-MIDDLE (single model) Anonymous	86.71	94.55

평가 항목

[아래 5개 태스크를 평가 항목으로 선정]

- 의미역 인식(Semantic Role Labeling): 문장 내에서 술어에 의해 기술되는 사건에 대한 개체들의 역할을 인식
- 기계 독해(Machine Reading Comprehension): 주어진 단락에서 질문에 요구하는 정답을 찾기
- 단락 순위화(Passage Ranking): 검색결과 집합에서 질문에 맞는 정답이 들어있는 단락 순위화
- 문장 유사도 추론(Natural Language Inference): 2개 문장 간 의미가 동일한 지 여부를 분류
- 문서 주제분류: 문서의 주제를 기정의된 54개의 클래스 중 하나로 분류

평가 결과

구분	의미역인식	기계독해	단락순위화	문장유사도추론	문서주제분류
평가데이터 및 규칙	Korean Propbank, 학습: 19,302 문장 평가: 3,773 문장	KorQuAD 데이터, 학습: 60,406건 평가: 5,773건 (dev-set)	학습: 45,521 질문 평가: 1,000 질문 (일문당 평균 8.7개 단락)	학습: 10,874문장쌍 평가: 1,209문장쌍 (이진 분류체계: 유사, 무관)	학습: 9,301건 평가: 1,035건 (54개 분류체계)
평가 방법	F1 ^[2]	Exact Match ^[3] / F1	Precision@Top1	Accuracy	Accuracy
(Google) Word Piece ^[4] 기반 한국어 언어모델	81.85%	80.82% / 90.68% (정답 경계 구분을 위해 후처리 수행)	66.3%	79.4%	91.1%
(엑소브레인) Word Piece 기반 한국어 언어모델	85.10%	80.70% / 91.94% (정답 경계 구분을 위해 후처리 수행)	70.5%	82.7%	93.4%
(엑소브레인) 형태소 기반 한국어 언어모델	85.77%	86.40% / 94.18%	73.7%	83.4%	93.7%

^[2] F1: 정답률(Precision) 시스템이 정답이 정답인 비율과 재현률(Recall) 실제 정답을 시스템이 맞춘 비율의 조화평균
^[3] Exact Match: 시스템이 제시된 질문과 정답이 완전히 일치하는 비율
^[4] Word Piece: 하나의 단어를 내부 단위(Subword Unit)들로 분리하는 단어 분의 방법



※ KorBERT 활용 논문 (서강대, 국립국어원 주관 국어정보처리시스템 경진대회 대상 수상)

4. 실험

본 논문에서는 ETRI[10]에서 사전 학습한 BERT base 모델을 사용하였다. BERT를 사용하기 위한 형태소 분석으로는 ETRI 형태소 분석기를 사용하였다. ELMo는 4GB의 대용량 뉴스 데이터를 이용하여 사전 학습하였다. 의존 구

표 1. 사전 학습 데이터에 따른 성능 비교(% , dev)

Model	EM	F1
BERT-google multi-lingual (baseline)	77.69	89.98
BERT-ETRI	84.82	92.74

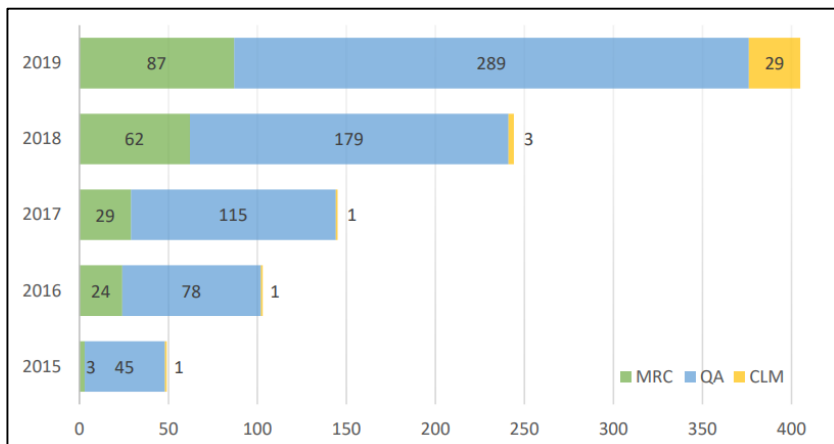
표 1. 상호참조해결 모델 별 성능 비교(자동 분석 형태소 이용)

Model	DEV			
	MUC	B ³	CEAF _F	CoNLL F1
규칙기반 [5]	44.25	45.21	50.91	46.79
Ptr-net [7]	59.07	55.16	53.58	55.94
Ptr-net + char + ELMo [7]	60.63	56.79	55.46	57.63
BERT-coref _{find} (ours)	49.62	47.79	46.29	47.90
BERT-coref (ours)	70.23	68.01	65.66	67.97
TEST				
Model	MUC	B ³	CEAF _F	CoNLL F1
Ptr-net + char + ELMo [7]	52.70	48.93	46.82	49.48
BERT-coref (ours)	65.98	64.03	63.64	64.55

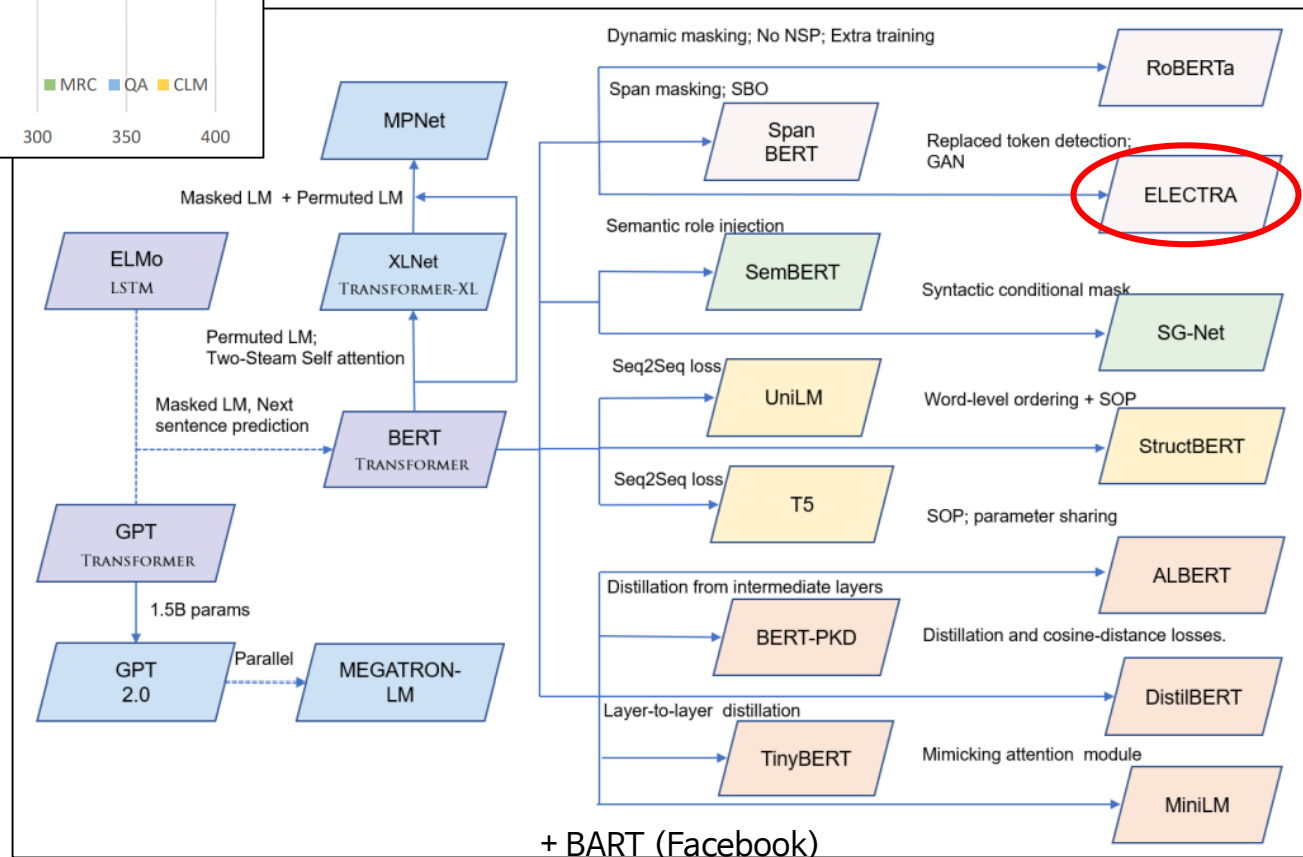
(2) 딥러닝 언어모델 (KorBERT)



• 최근 연구 동향



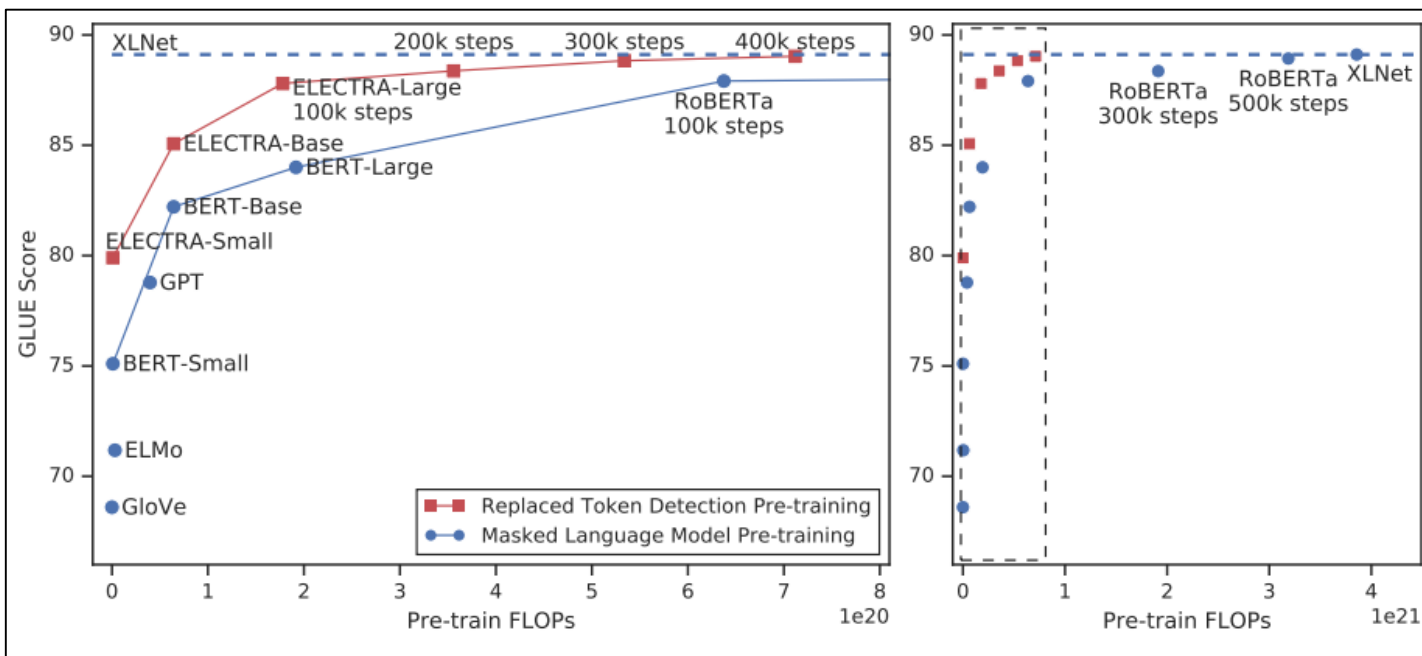
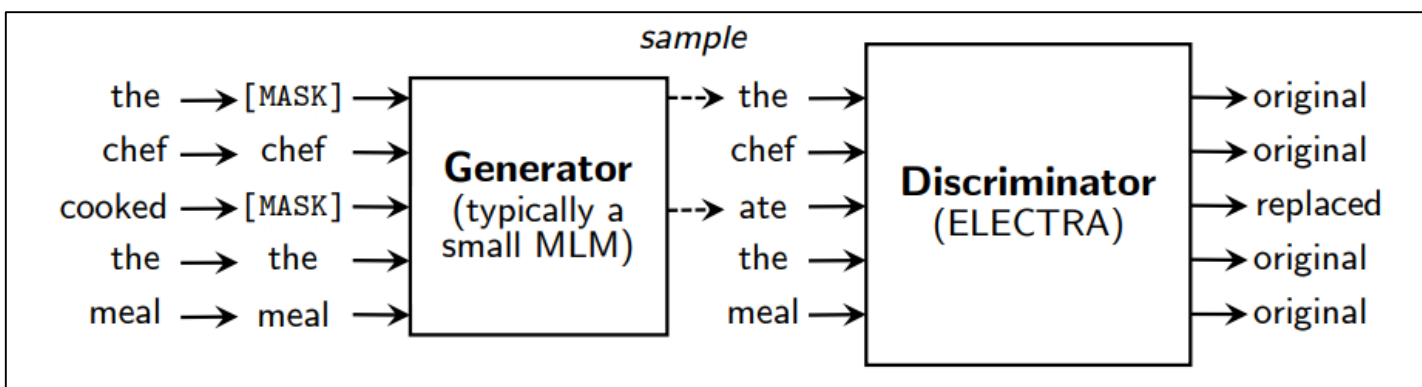
※ 출처: <https://arxiv.org/pdf/2005.06249.pdf>



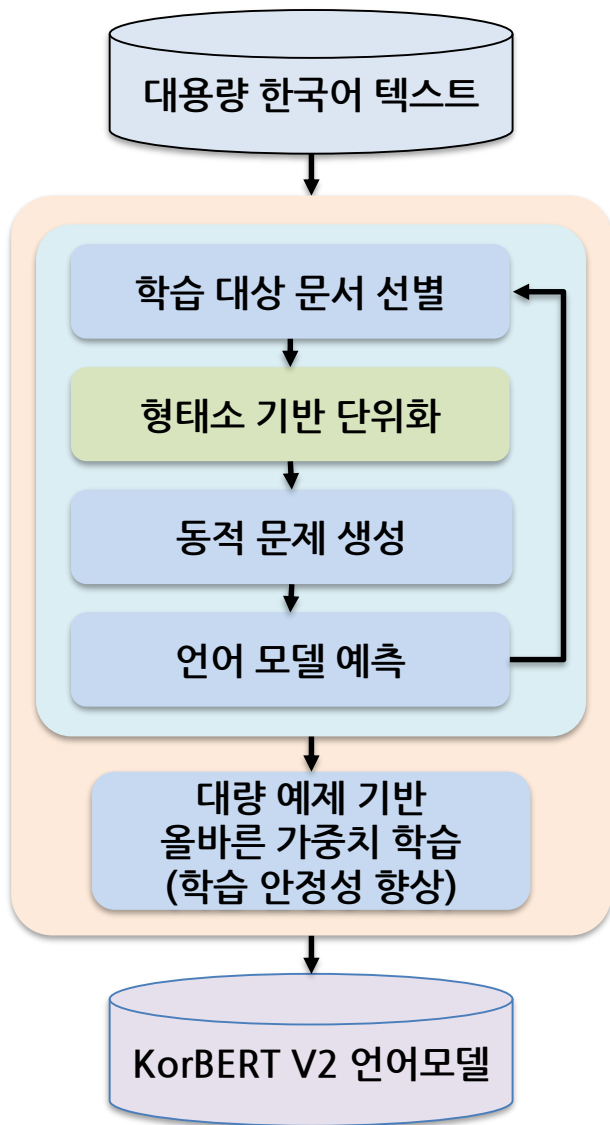
(2) 딥러닝 언어모델 (KorBERT)



- 연구동향
 - ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators



(2) 딥러닝 언어모델 (KorBERT)



항목	AS-IS	TO-BE
학습 데이터	문어 학습데이터 - 문어 23GB (47억 형태소)	<u>문어+구어 학습데이터</u> - 문어+구어 32GB (76억 형태소)
학습 알고리즘	기본: Masked LM (with 다양한 LM 학습 기법)	기본: 생성-분류 구조 <u>장기: 멀티-태스크</u> (for general task learning)
적용 태스크	문어 도메인 (언어이해/기계독해 /문서분류 등)	문어/구어 도메인 (언어이해/기계독해/ 문서분류/대화처리 등)
서비스 모델	Base 모델	Base & 경량화 모델 (Knowledge Distillation)

발표 순서

- I 엑소브레인 과제 개요
- II 심층 언어분석 기술
- III 딥러닝 언어모델 기술
- IV 심층 질의응답 기술**



- 기계독해 vs. 질의응답

- 기계독해

- 주어진 질문과 단락에 대해, 정답(또는 no answer) 추론 문제

- 질의응답

- 주어진 질문에 대해, 전체 문서 집합에서 정답(또는 no answer) 추론 문제

※ (SQuAD v1.1 대상) 기계독해 vs 질의응답 시, 정답 인식 성능 평가

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

SQuAD1.1 Leaderboard

Here are the ExactMatch (EM) and F1 scores evaluated on the test set of SQuAD v1.1.

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar et al. '16)	82.304	91.221
1	LUKE (single model) Studio Ousia & NAIST & RIKEN AIP	90.202	95.379

Models	F1	EM
multi-passage (Wang et al., 2019b)	60.9	53.0
ORQA (Lee et al., 2019)	-	20.2
BM25+BERT (Lee et al., 2019)	-	33.2
Weaver (Raison et al., 2018)	-	42.3
RE ³ (Hu et al., 2019)	50.2	41.9
MUPPET (Feldman & El-Yaniv, 2019)	46.2	39.3
BERTserini (Yang et al., 2019)	46.1	38.6
DENSPI-hybrid (Seo et al., 2019)	44.4	36.2
MINIMAL (Min et al., 2018)	42.5	34.7
Multi-step Reasoner (Das et al., 2019)	39.2	31.9
Paragraph Ranker (Lee et al., 2018)	-	30.2
R ³ (Wang et al., 2018a)	37.5	29.1
DrQA (Chen et al., 2017)	-	29.3
Ours	63.8	56.5

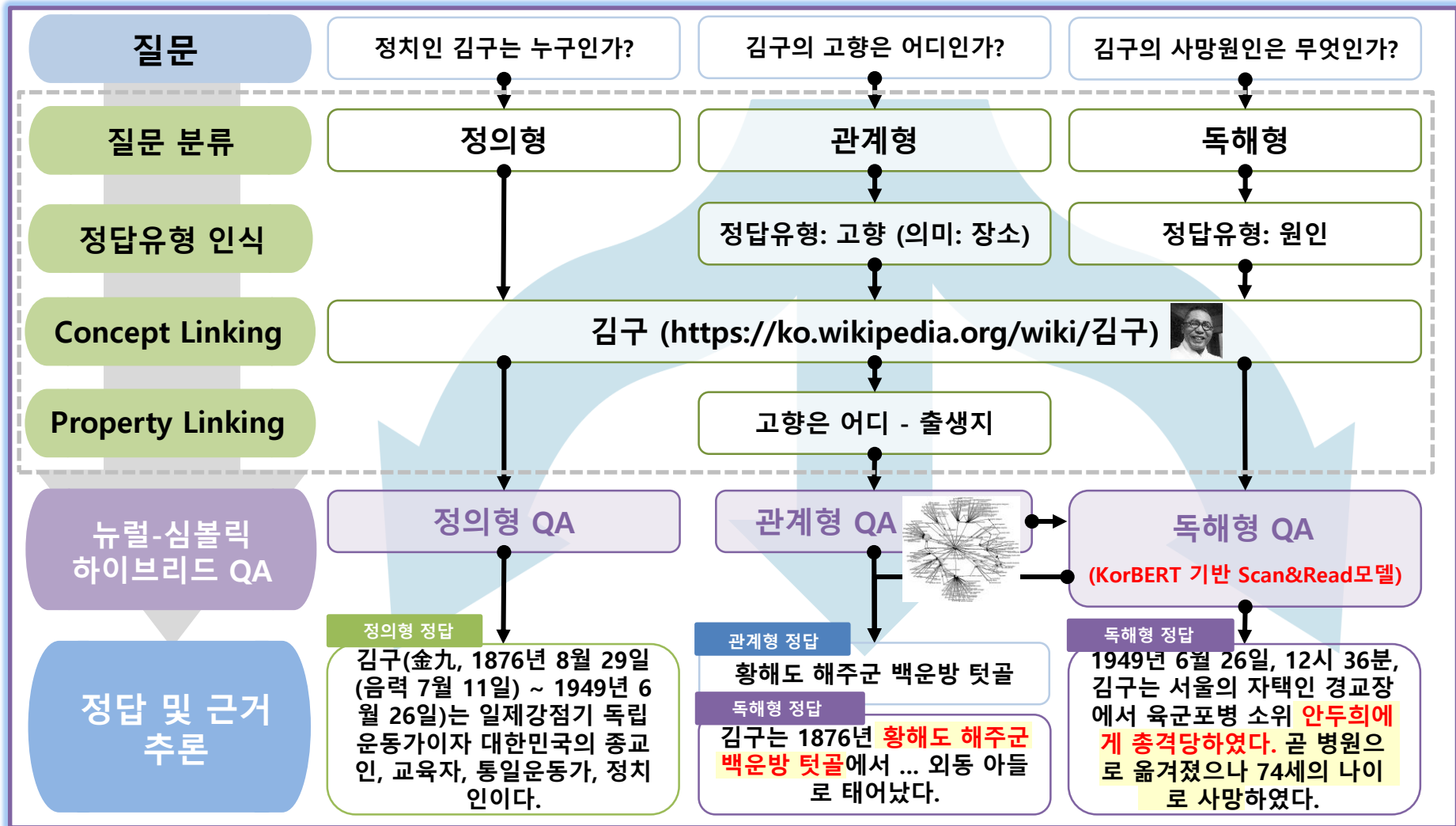
Table 3: **SQuAD Open results**: we report F1 and EM scores on the test set of SQuAD Open, following previous work.

* <https://arxiv.org/pdf/1911.10470.pdf>
Learning to Retrieve Reasoning Paths
over Wikipedia Graph for Question Answering

(3) 뉴럴-심볼릭 하이브리드 심층 질의응답 (WiseQA)



문제유형 별로 뉴럴 또는 심볼릭 QA방법을 Planning하여 최적정답 생성



(3) 뉴럴-심볼릭 하이브리드 심층 질의응답 (WiseQA)



검색된 다수 단락으로부터 **상용화 수준(속도/성능)의 정답 독해** 기술 개발

1. Scan 모델 : 정답 추출 확률이 높은 순으로 단락 순위화, 정답 추출 확률이 낮은 단락 제거
2. Read 모델 : 질문과 단락으로부터 단답형 또는 서술형 정답의 경계(시작위치~끝위치) 인식

* 구글도 BERT 모델을 적용하여 검색결과를 재순위화한 서비스 발표 ('19.10)

[KorBERT-Scan 모델]

1) 학습 목표

- P(정답존재여부 | 질문, 단락)

2) 학습 데이터

- 정답/오답 학습데이터 자동 생성

3) 결과 활용

- 정답 추출 확률이 높은 순으로 재순위화
- 정답 추출 확률이 낮은 단락 제거

4) 예시

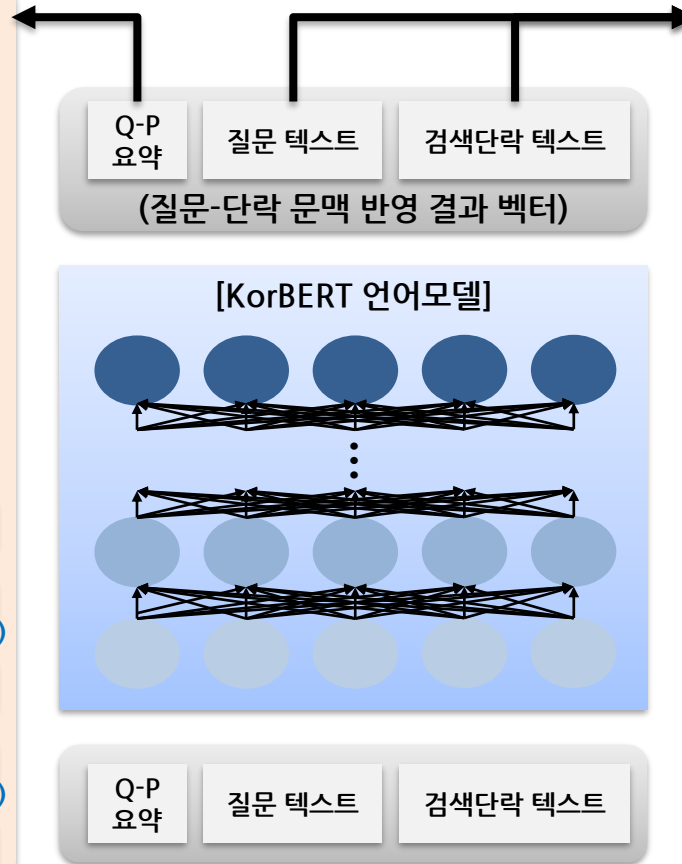
질문: 헌법상 대통령의 임기는 얼마인가?

검색단락#1: 대통령의 임기가 만료된 때에는...
(no-answer prediction: 제거)

검색단락#2: 원장은 대통령이 임명하고, 그 임기는 ...

검색단락#3: 대통령의 임기연장 또는 중임변경을 ...
(no-answer prediction: 제거)

검색단락#4: 대통령의 임기는 5년으로 하며, 중임 ...



[KorBERT-Read 모델]

1) 학습 목표

- P(정답시작, 정답끝 | 질문, 단락)

2) 학습 데이터

- 질문 / 단락 / 정답경계 학습 데이터

3) 결과 활용

- 단락 내 올바른 정답 경계 인식

4) 예시

질문: 헌법상 대통령의 임기는 얼마인가?

<1순위 정답> 5년

근거: 대통령의 임기는 5년으로 하며, 중임할 수 없다.

출처: 대한민국헌법 | 제4장 정부 / 제1절 대통령 ...

<2순위 정답> 4년

근거: ... 그 임기는 4년으로 하며, ...

출처: 대한민국헌법 | ... 제4관 감사원 ...



• (Scan 모델) 질문-단락 관련성 학습 태스크

– Scan 모델 입력/출력 예 (질문: **대통령의 임기는 얼마인가?**)

검색 순위	Scan 모델 순위	Scan 모델 점수	기계독해 정답	근거 단락
1	4	0.0002	70일 내지 40일전	① 대통령의 임기 가 만료되는 때에는 임기 만료 70일 내지 40일 전에 후임자를 선거한다.
2	2	0.0011	4년	② 원장은 국회의 동의를 얻어 대통령 이 임명하고, 그 임기 는 4년으로 하며, 1차에 한하여 중임할 수 있다.
3	3	0.0004	중임변경을 위한 헌법개정	② 대통령의 임기 연장 또는 중임변경을 위한 헌법개정은 그 헌법개정 제안 당시의 대통령 에 대하여는 효력 ...
4	1	0.9997	5년	대통령의 임기 는 5년으로 하며, 중임할 수 없다.

– Top1 재현율 82.60% 달성 (검색대비 +5.67%)

- 학습데이터 자동생성 기술을 이용한 839,381건 학습데이터 생성
- 검색 단락 필터링 비율 67.71% / 정확률 85.65% 달성


이진 재현율	Top1	Top2	Top3
시맨틱 검색	76.93%	81.93%	84.07%
구글 WordPiece 모델	67.67%	81.73%	86.67%
KorBERT WordPiece 모델	74.53%	84.47%	87.20%
KorBERT 형태소 모델	82.60% (+5.67%)	86.67%	88.47%

(3) [활용사례] 법률분야 심층질의응답 시스템



국내외 법 조항에 관해서 단답형/서술형 질문에 정답을 제공하는 질의응답 기술
- 법령 질문 중 50%는 정답길이가 10글자 이상인 고난도 서술형 문제 -

- 국회도서관의 입법관련 입안/심사 조사업무를 효율적으로 지원하기 위한 법령 QA 서비스
 - 국내 법령(1,456건) 및 35개국 헌법(번역본) 대상 단답형/서술형 QA 서비스
 - 국회도서관 / 국가과학기술연구회 '20년도에 기술도입 (진행 중)
- 법령분야 질의응답 오픈API 서비스 ('19.12.20)



법률 질의응답 시스템

언어를 이해하고, 지식을 학습하여, 법률지식을 제공하는 질의응답 시스템
(본 시스템은 Chrome 브라우저 사용을 권장합니다.)

◆ 정의형 질문

1. 김영란법이 뭐야?
2. 공인중개사법에서 중개의 의미는?
3. 대한민국 영토?
4. 정치자금법의 계정 목적은 무엇인가?
5. 공공기관의 뜻은?

◆ 단답형 질문


1. 헌법상 국회의원의 임기는?
2. 대법원장의 임기는 몇년인가?
3. 대한민국의 주권은 누구에게 있나?
4. 국회의 도서 및 입법자료에 관한 업무 처리하기 위한 조직은?
5. 대한민국헌법은 언제 제정되었어?

◆ 서술형 질문


1. 국회 상임위원회는 예산안과 결산이 회부되면 무엇을 하는가?
2. 대통령령이 제정·개정 또는 폐지되었을 경우 해야하는 일은?
3. 국회에서 의원이 발언을 하기 위한 방법은?
4. 국무총리가 사교로 직무를 수행할 수 없는 경우 정부조직법상 누가 직무를 대행하나요?
5. 국회에서 긴급한 경우 본회의 개의 방법은?

◆ 사용자 질문

1. 공인중개사법에서 중개의 의미는?
2. 대한민국의 국기는?
3. 사람의 생명 또는 신체에 중대한 위해를 발생하게 할 우려가 있는 수술, 수혈, 전신마취에 따른 설명, 동의 및 고지의 방법·절차 등 필요한 사항은 어떻게 정하나요?
4. 대한민국헌법 전문에 따른 때 국민이 계승하여야 하는 법률은 무엇인가?
5. 대법관의 임기는?



대한민국의 주권은 누구에게 있는가?




사용자의견

응답결과


① 국민

대한민국 헌법 | 제1장 총강 | 제1조

... ② 대한민국의 주권은 국민에게 있고, 모든 권력은 국민으로부터 나온다.



대통령령이 제정·개정 또는 폐지되었을 경우 해야하는 일은?



사용자의견

응답결과

① 10일 이내에 이를 국회 소관 상임위원회에 제출

국회법 | 제6장 회의 / 제3절 의사와 수정 | 제98-2조 대통령령 등의 제출 등

① 중앙행정기관의 장은 법률에서 위임한 사항이나 법률을 집행하기 위하여 필요한 사항을 규정하는 대통령령·총리령·부령·훈령·예규·고시 등이 제정·개정 또는 폐지되었을 때에는 10일 이내에 이를 국회 소관 상임위원회에 제출하여야 한다. 다만, 대통령령의 경우에는 입법예고를 할 때(입법예고를 생략하는 경우에는 법제처장에게 심사를 요청할 때를 말한다)에도 그 입법예고안을 10일 이내에 제출하여야 한다.

② 중앙행정기관의 장은 제1항의 기간 이내에 제출하지 못한 경우에는 그 이유를 소관 상임위원회에 통지하여야 한다.

③ 상임위원회는 위원회 또는 상설소위원회를 정기적으로 개최하여 그 소관 중앙행정기관이 제출한 대통령령·총리령 및 부령(이하 이 조에서 "대통령령등"이라 한다)의 법률 위반 여부 등을 검토하여 법률의 취지 또는 내용에 합치되지 아니한다고 판단되는 경우에는 소관 중앙행정기관의 장에게 그 내용을 통보할 수 있다. 이 경우 중앙행정기관의 장은 통보받은 내용에 대한 처리 계획과 그 결과를 지체 없이 소관 상임위원회에 보고하여야 한다.

④ 전문위원은 제3항에 따른 대통령령등을 검토하여 그 결과를 해당 위원회 위원에게 제공한다.

<법령 질의응답 서비스>

(3) [활용사례] 한컴오피스 2020 지식검색 서비스



- 위키백과QA의 한컴오피스 2020 신규버전('19.10)에 탑재
 - 한컴 말랑말랑 서비스를 통한 오피스톡 및 지식검색 서비스 공개



- 클라우드 환경 오피스 문서 대상 QA 시스템(ODWiseQA) 개발 (진행 중)





- 연구동향

- Evidence Detection in Reading Comprehension

- HotpotQA: 명시적 근거 문장 데이터 기반 지도학습

Paragraph A, Return to Olympus:

[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:

[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] *The band was active from 1987 to 1990.* [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] *The album was finally released a few months later.*

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

A: Malfunkshun

Supporting facts: 1, 2, 4, 6, 7

Figure 1: An example of the multi-hop questions in HOTPOTQA. We also highlight the supporting facts in *blue italics*, which are also part of the dataset.

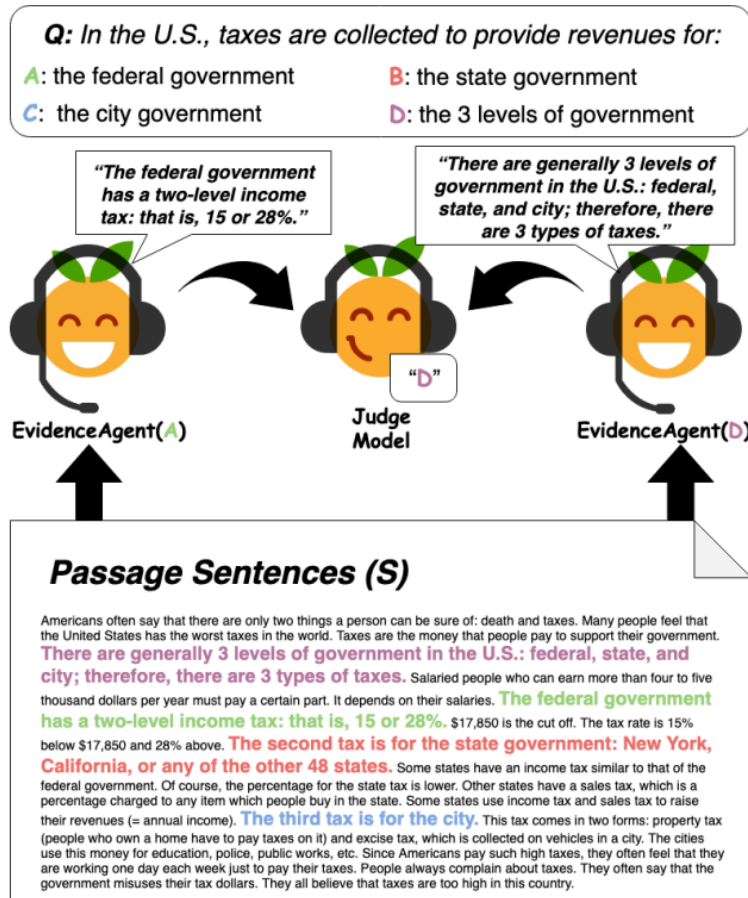
Leaderboard (Fullwiki Setting)

In the fullwiki setting, a question-answering system must find the answer to a question in the scope of the entire Wikipedia. Similar to in the distractor setting, systems are evaluated on the accuracy of their answers (Ans) and the quality of the supporting facts they use to justify them (Sup).

	Model	Code	Ans		Sup		Joint	
			EM	F ₁	EM	F ₁	EM	F ₁
1	Step-by-Step Retriever (single model)		62.95	75.43	54.61	80.00	60.36	66.22
2	HopRetriever-V1 (single model)		60.83	73.93	53.07	79.26	58.00	63.91
3	DDRQA (single model)		62.53	75.91	51.01	78.86	56.04	63.88
4	DR model large (single model)		62.01	75.32	49.88	77.77	55.44	62.95
5	HGN-albert + SemanticRetrievalMRS IR (single model)		59.74	71.41	51.03	77.37	57.92	62.26
6	Robustly Fine-tuned Graph-based Recurrent Retriever (single model)		60.04	72.96	49.08	76.41	55.35	61.18
7	HGN-large + SemanticRetrievalMRS IR (single model)		57.85	69.93	51.01	76.82	53.71	60.74



- 연구동향
 - Evidence Detection in Reading Comprehension
 - (2020.09) Finding Generalizable Evidence by Learning to Convince Q&A Models
 - Judge(기계독해, 인간 등)의 올바른 정답 추론 확률을 최대로 하는 부분 근거 문장 인식



$$\arg \max_{E(i) \subset S} \log p_{\phi}(i|E(i), Q, A)$$

※ 출처: Learning to find evidence (Kyunghyun Cho, NYU)

Summary

Taxes are the money that people pay to support their government. There are generally three levels of government in the United States; federal, state and city. The federal government has a two-level income tax; that is, 15 or 28 percent. Some states use income tax and sales tax to raise their revenues. The cities use this money for education, police, public works, etc.

(3) 설명 가능한 심층 질의응답 기술



연구 목표

- **복수의 연관된 근거들 사이 인과관계** 기반 정답 추론
 - 현재 기계독해 기술 = 단일 단락 내 질문의 정답 및 근거가 포함되어 있어야 하는 한계
- **정답/근거를 설명 가능한 질의응답**

연구 현황 및 계획

- 명시적 복합 근거를 포함한 기계독해 학습데이터 구축
- **질문-검색 단락 기반 개별 근거 인식** 모델 개발
- **복합 근거 기반 정답 추론 및 순위화** 모델 개발

(국회도서관 실증테스트 질문) 변리사에 의한 손해는 배상 받을 수 있나요?

AS-IS

변리사법

제6조의3(특허법인의 설립) ① 변리사는 업무를 조직적·전문적으로 수행하기 위하여 대통령령으로 정하는 바에 따라 **3명 이상의 변리사를 구성원으로 하는 특허법인을 설립할 수 있다.** <개정 2013. 7. 30., 2016. 1. 27.> ...

변리사법

제6조의18(특허법인(유한)의 손해배상준비금 등) ① **특허법인(유한)은 업무를 수행하다가 위임인에게 손해를 입힌 경우 그 손해에 대한 배상책임을 보장하기 위하여 대통령령으로 정하는 바에 따라 사업연도마다 손해배상준비금을 적립하거나 손해배상책임보험에 가입하여야 한다.** ...

단일 단락 내 질문-정답근거 미존재
→ 검색 결과 제시

TO-BE

변리사법

제6조의18(특허법인(유한)의 손해배상준비금 등) ① **특허법인(유한)은 업무를 수행하다가 위임인에게 손해를 입힌 경우 그 손해에 대한 배상책임을 보장하기 위하여 대통령령으로 정하는 바에 따라 사업연도마다 손해배상준비금을 적립하거나 손해배상책임보험에 가입하여야 한다.** ...

변리사법

제6조의3(특허법인의 설립) ① 변리사는 업무를 조직적·전문적으로 수행하기 위하여 대통령령으로 정하는 바에 따라 **3명 이상의 변리사를 구성원으로 하는 특허법인을 설립할 수 있다.** <개정 2013. 7. 30., 2016. 1. 27.> ...

질문-검색 단락 기반 개별 근거 인식 모델

근거#1: **특허법인(유한)은 업무를 수행하다가 위임인에게 손해를 입힌 경우 ...**

근거#2: **3명 이상의 변리사를 구성원으로 하는 특허법인을 설립할 수 ...**

복합 근거 기반 정답 추론 및 순위화 모델

정답: YES

- 특허법인(유한)은 업무를 수행하다가 위임인에게 손해를 입힌 경우 그 손해에 대한 배상책임을 진다.
- 특허법인은 대통령령으로 정하는 바에 따라 3명 이상의 변리사를 구성원으로 설립할 수 있다.



감사합니다