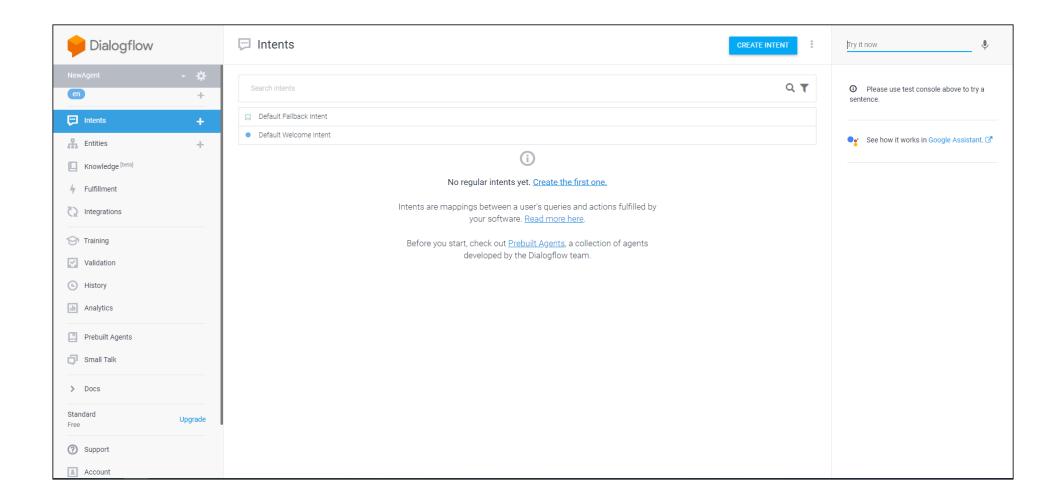
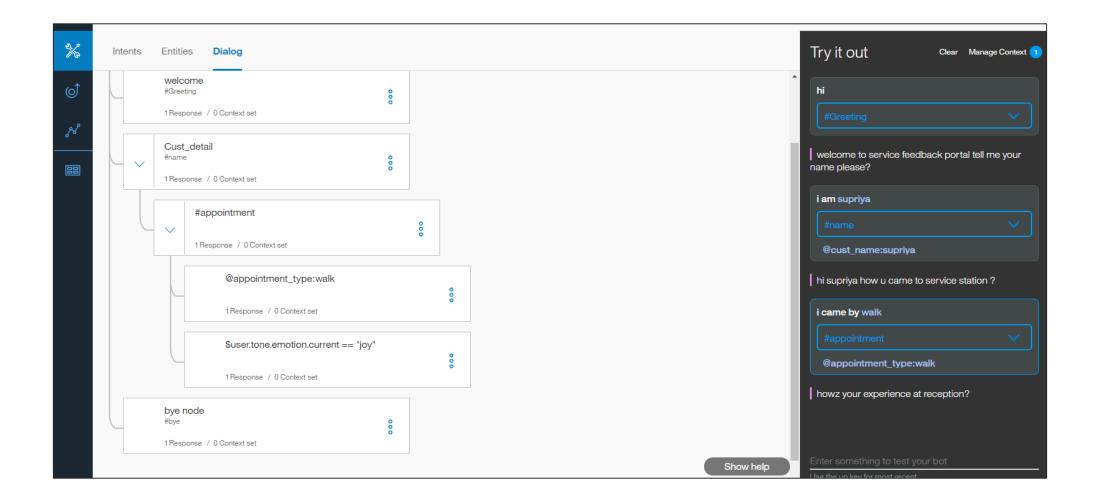
MAUM SDS

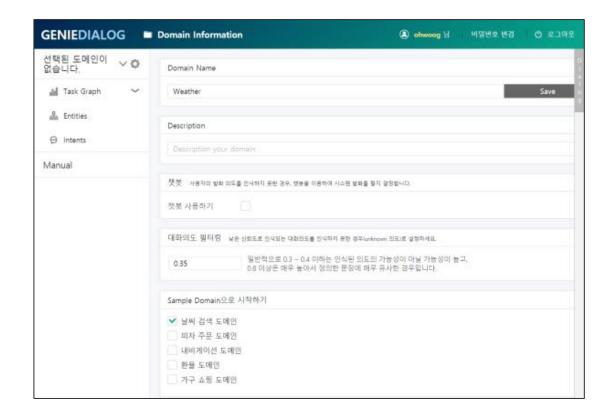
Week_02

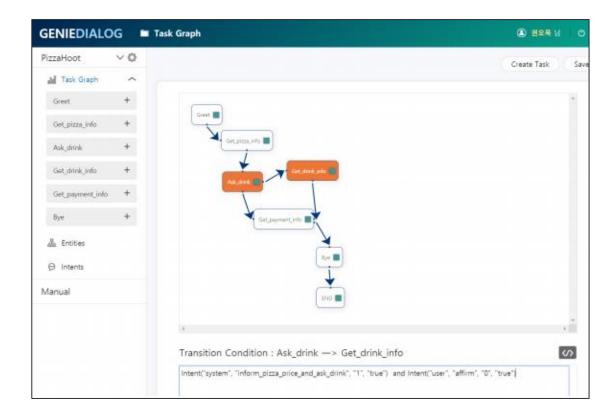
마인즈랩 챗봇 소개

챗봇은 만들기 쉽다. 챗봇은 유지하기 어렵다









하이브리드 기반 접근 방식

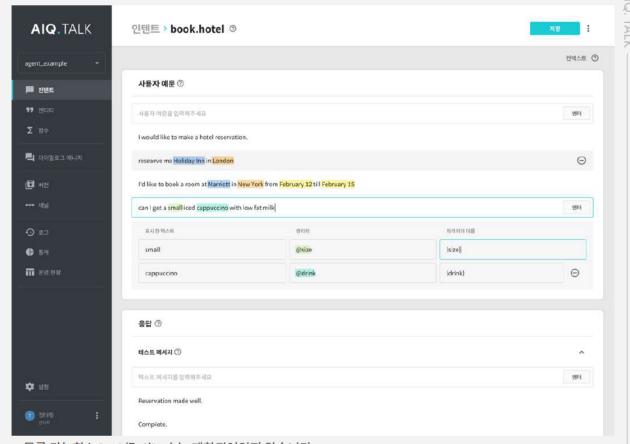
Intent 학습 문장

사용자 발화 예문

Intent 학습 문장(사용자 발화 예문)을 입력하고, 입력된 문장을 기반으로 Intent 분류 모델을 학습

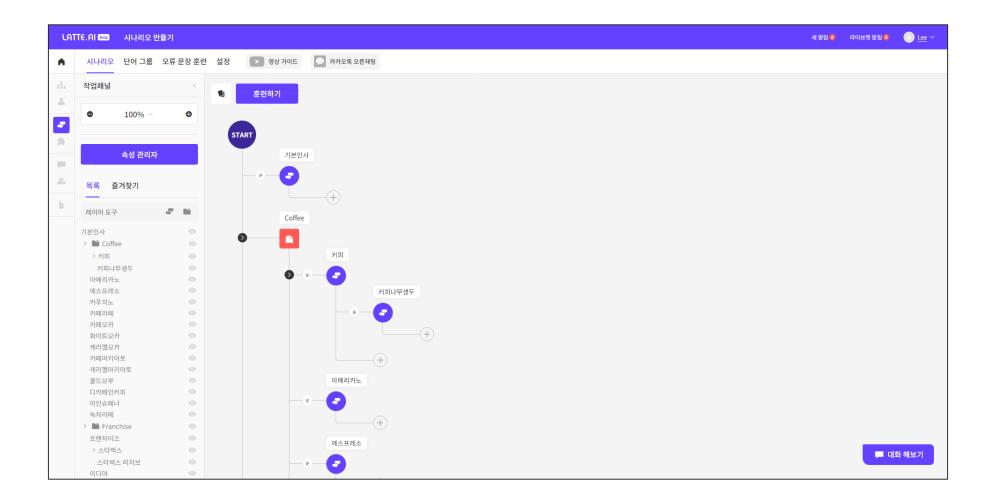
엔티티 어노테이션(Entity Annotation)

Intent 학습 문장은 실 발화 예상 문장 뿐만 아니라, 대표 엔티티가 포함된 형태의 문장도 함께 구성



- 등록 가능한 Intent/Entity 수는 제한되어있지 않습니다.
- 단, Intent/Entity 수가 매우 많을 경우에는 성능 및 학습 속도에 영향을 끼칠 수 있습니다.





챗봇은 만들기 쉽다. 챗봇은 유지하기 어렵다

Q1

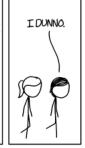
만들기 쉬운데, 왜 챗봇에 대한 접근성은 저조한가?

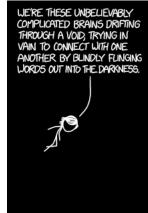
이상







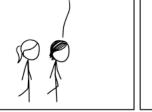




EVERY CHOICE OF PHRASING AND SPEILING AND TONE AND TIMING CARRIES COUNTLESS SIGNALS AND CONTEXTS AND SUBTEXTS AND MORE,

AND EVERY LISTENER INTERPRETS THOSE SIGNALS IN THEIR OWN WAY.

LANGUAGE ISN'T A FÒRMAL SYSTEM. LANGUAGE IS GLORIOUS CHAOG.



YOU CAN NEVER KNOW FOR SURE WHAT ANY WORDS WILL MEAN TO ANYONE.

ALL YOU CAN DO IS TRY TO GET BEITER AT GUESGING HOW YOUR WORDS AFFECT PEOPLE, SO YOU CAN HAVE A CHANCE OF FINDING THE ONES THAT WILL MAKE THEN FEEL SOMETHING LIKE WHAT YOU WANT THEN TO FEEL.

EVERYTHING ELSE IS POINTLESS.



I ASSUME YOU'RE GIVING ME TIPS ON HOW YOU INTERPRET WORDS BECAUSE YOU WANT ME TO FEEL LESS ALONE.

IF 50, THEN THANK YOU. THAT MEANS A LOT.



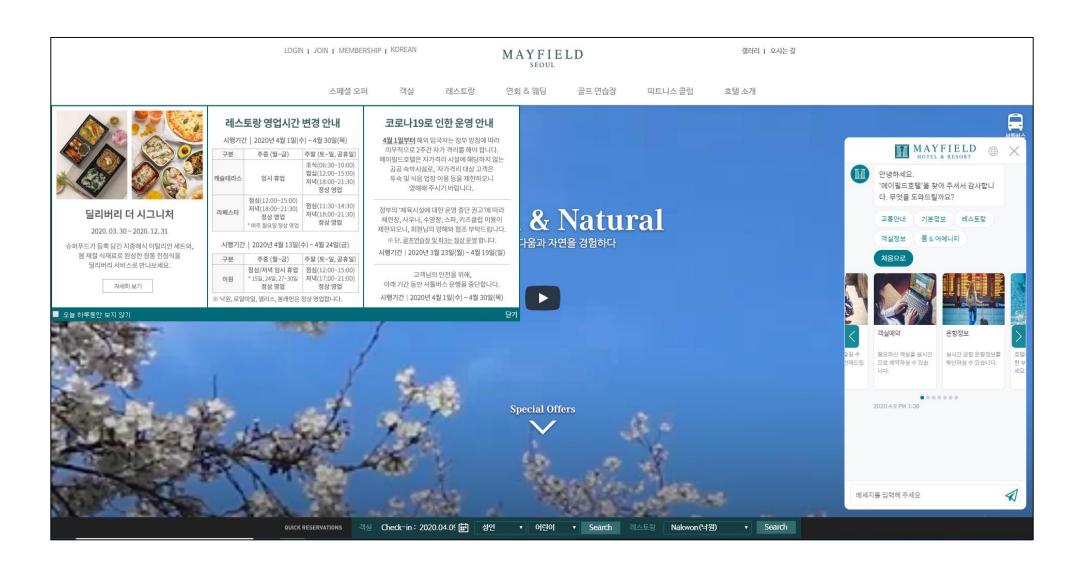
BUT IF YOU'RE JUST RUNNING MY SENTENCES PAST SOME MENTAL CHECKLIST SO YOU CAN SHOW OFF HOW WELL YOU KNOW IT,



THEN I COULD CARE LESS.



현실



Answer

쉽게 재설계

Q2 유지관리 어떻게 해야 쉬운 방법이 있나?

Q2

- 1. 대량의 데이터를 처리
- 2. 단순 업무지만 숙련된 작업자 필요
- 3. 현재 인공지능은 단순작업 중심

Answer

- 1. 대량의 데이터를 효율적으로 관리해야 함
- 2. 비숙련자도 쉽게 이해할 수 있어야 함
- 3. 고도화된 인공지능을 적용해야 함

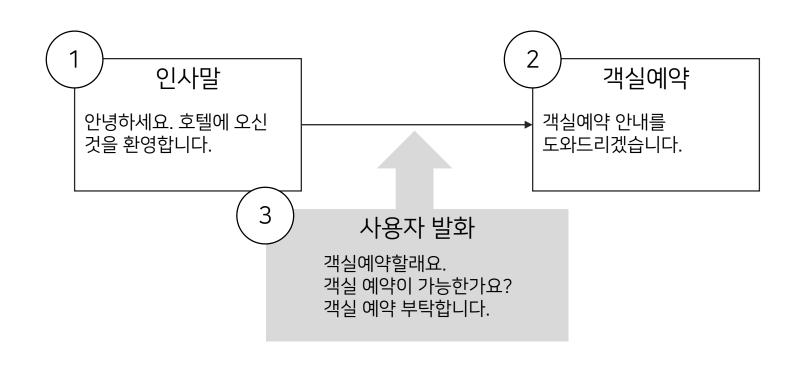
챗봇을 쉽게 만들자

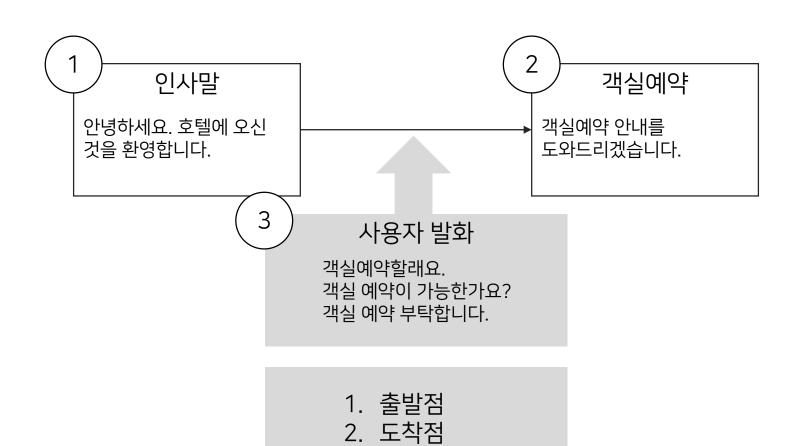




사용자 발화

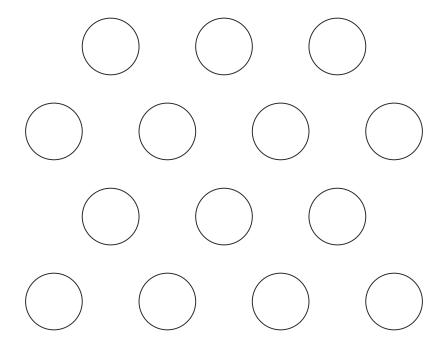
객실예약할래요. 객실 예약이 가능한가요? 객실 예약 부탁합니다.



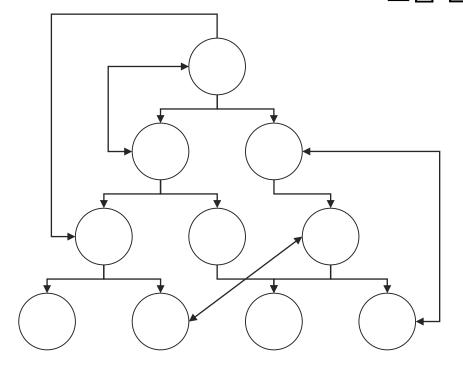


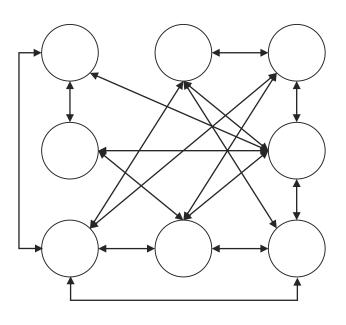
3. 사용자발화

답변 만들기 = Intent 제작

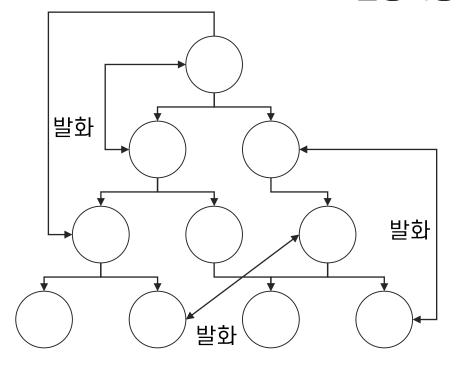


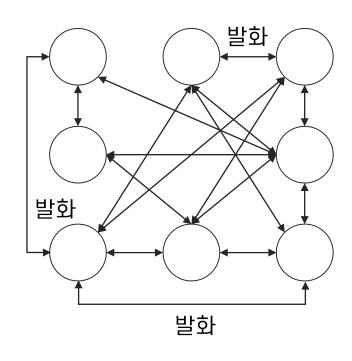
흐름 만들기 = Flow 제작





인공지능 만들기 = BERT 제작





답변 만들기 Intent 제작	흐름 만들기 Flow 제작	인공지능 만들기 BERT 제작
답변.xlsx	로직.xlsx	학습.xlsx

답변 만들기 Intent 제작		흐름 만들기 Flow 제작		인공지능 만들기 BERT 제작
답변.¥lsx	+	로직.xlsx	+	학습.xlsx



ITC와 NER 소개

ITC(Intent Classifier) 의도분류

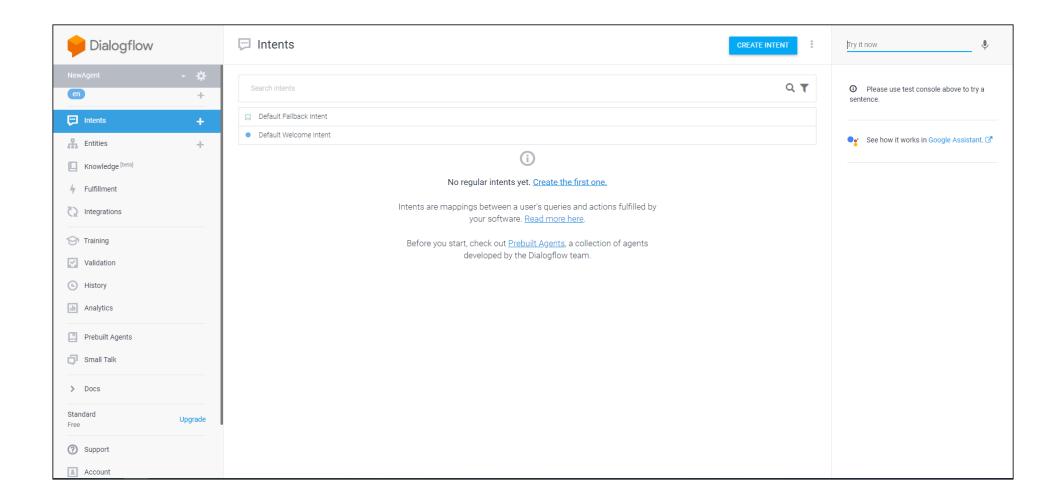
문장의 의미 또는 목적을 찾아낸다.

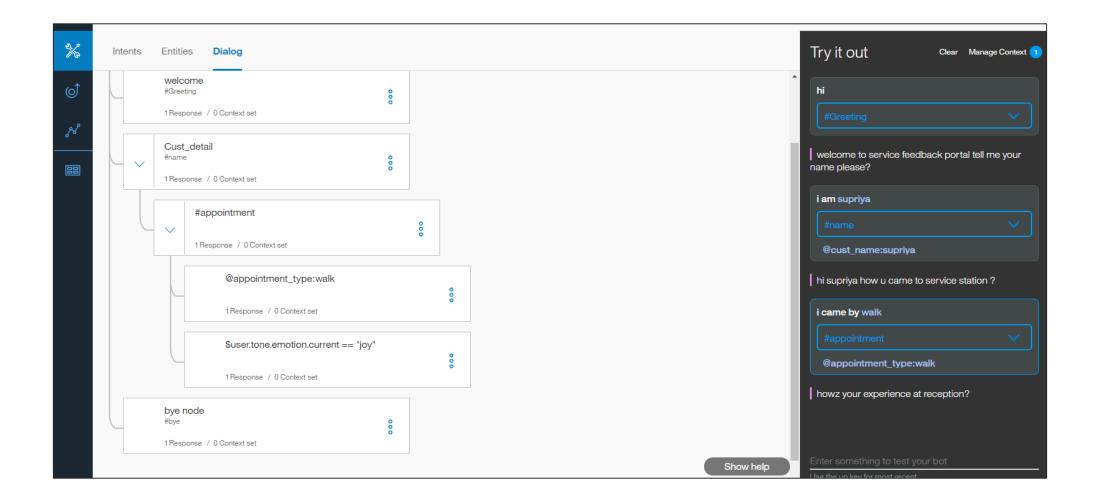
NER(Named Entity Recognition) 개체명 인식

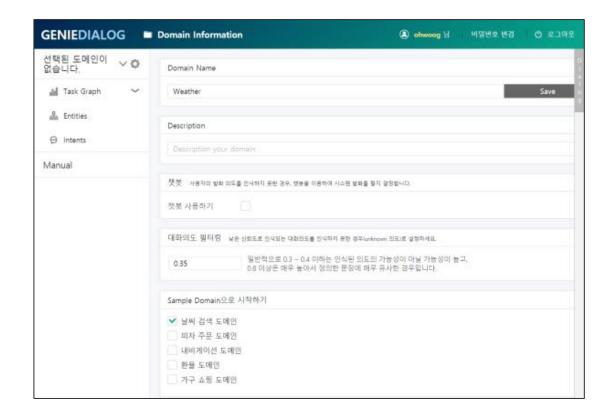
문장에서 특정한 종류의 단어를 찾아내는 Information Extraction을 의미한다. Entity 또는 Slotmapping 이라고도 불린다.

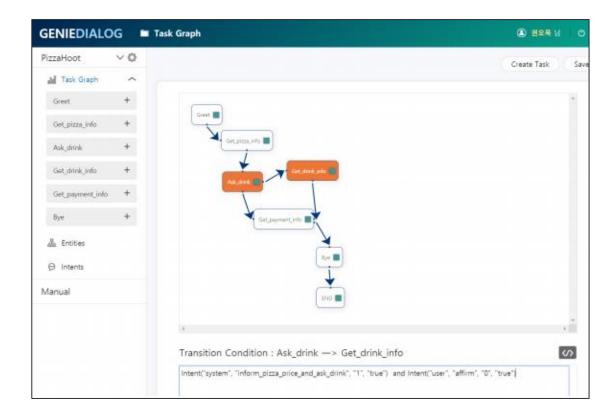
Input	모짜르트 클래식 틀어줘
	(Play Mozart's classical music)
Intent	play_music
Entities	genre: 클래식(classic),
	composer: 모짜르트(Mozart)

Table 1: An example of analyzing a chatbot input sentence.









하이브리드 기반 접근 방식

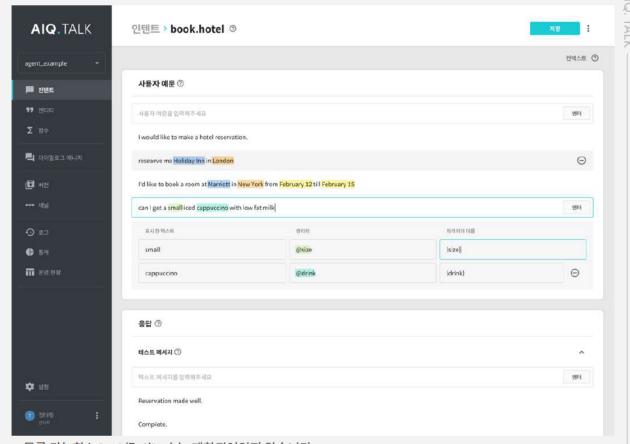
Intent 학습 문장

사용자 발화 예문

Intent 학습 문장(사용자 발화 예문)을 입력하고, 입력된 문장을 기반으로 Intent 분류 모델을 학습

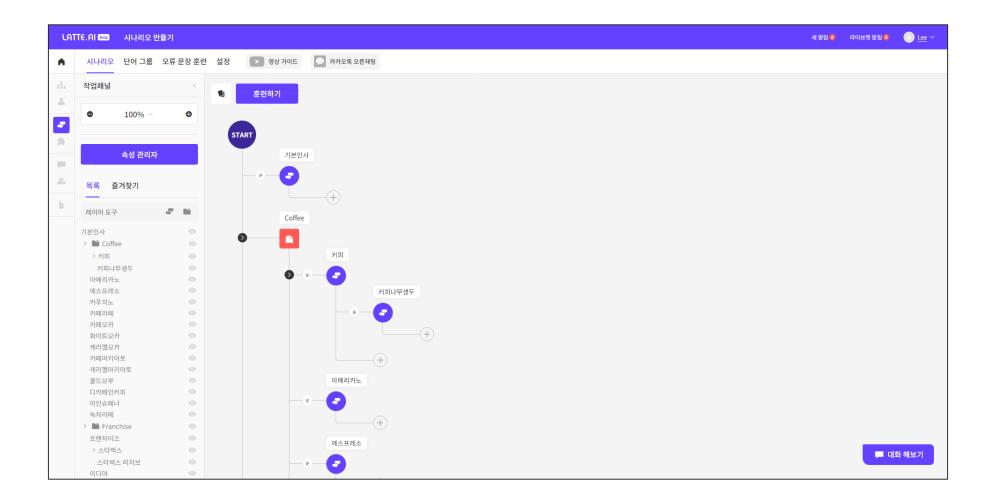
엔티티 어노테이션(Entity Annotation)

Intent 학습 문장은 실 발화 예상 문장 뿐만 아니라, 대표 엔티티가 포함된 형태의 문장도 함께 구성

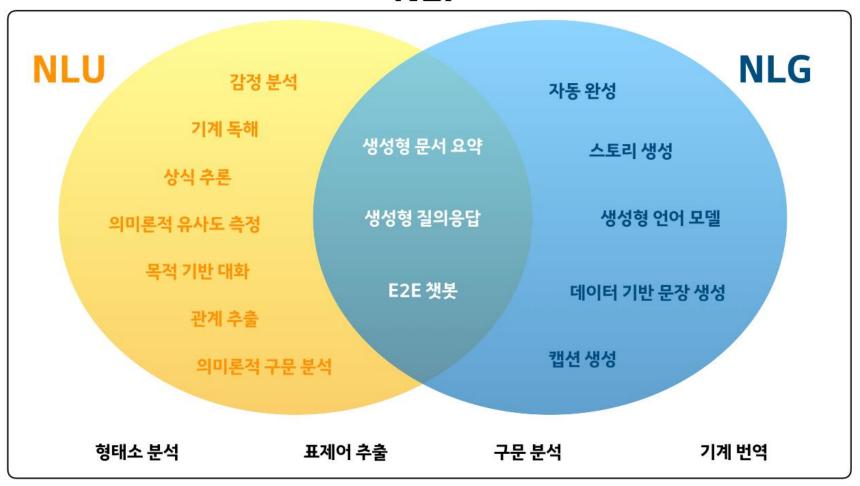


- 등록 가능한 Intent/Entity 수는 제한되어있지 않습니다.
- 단, Intent/Entity 수가 매우 많을 경우에는 성능 및 학습 속도에 영향을 끼칠 수 있습니다.









과제	벤치마크		예제	
	SST, IMDb,	입력	문장: 신만이 이 영화를 용서할수 있다	
	NSMC	출력	긍정 · <mark>부정</mark>	
유사도 예측 Similarity Prediction	STS, MRPC, QQP, PAWS-X	입력	문장1: 파이썬과 자바 중 뭐부터 배워야 하나요? 문장2: Java나 Python 중 하나를 배워야 한다면, 뭐부터 시작해야 할까요?	
		출력	<mark>유사</mark> · 비유사	
자연어 추론 Natural Language	SINLI, MINLI,	입력	전제: 회색 개가 숲에서 쓰러진 나무를 핥고 있다. 가설: 개가 밖에 있다.	
Inference		출력	<mark>참</mark> · 거짓 · 모름	
언어적 용인 가능성		입력	문장: 그는 한 번도 프랑스에 갔다.	
Linguistic Acceptability	CoLA	출력	가능 · <mark>불가능</mark>	
기계 독해 Reading Comprehension	MS MARCO, KorQuAD	입력	지문: 시카고 대학의 학자들은 경제학, 사회학, 법학 분석, 문학 비평, 신학, 행동주의 정치학 등 다양한 학문 발전에 중요한 역할을 담당해왔다. […] 이 대학은 미국최대 대학 출판사인 시카고 대학 프레스(University Press of Chicago Press)의 본거지이기도 하다. 오는 2020년 완공될 버락 오바마 대통령 센터에는 오바마 대통령 도서관과 오바마 재단 사무소가 세워질 예정이다.	
		출력	2020년	
의도 분류 Intent Classification	ATIS, SNIPS, AskUbuntu	입력	질문: 현재 위치 주변에 있는 맛집 두 곳만 알려줘	
		출력	장소 검색 (식당, 가까운, 평점 좋은, 2)	

https://www.kakaobrain.com/blog/118#ref_list_24

