

2021 국립국어원 인공지능 언어 능력 평가 대회

· 과제명: 동형이의어 구별

· 과제 제작: 국립국어원, 에스케이텔레콤

· 자료 설명

· 목표

“동형이의어 구별”은 두 개의 서로 다른 맥락에서 동일한 ‘형태’의 단어가 사용되었을 때 그 단어의 ‘의미’도 같은지 판별하는 과제입니다. 언어 모델이 맥락에 따른 단어의 의미를 구별할 수 있는지 평가하게 됩니다. 본 자료는 슈퍼글루(SuperGLUE)의 WiC(Words in Context, 맥락 기반 어휘)를 참조하여 제작하였습니다. 동사·형용사 자료는 국립국어원의 “2020년 어휘 의미 말뭉치 연구 분석 사업” 사업 결과물을 재구성하였으며 명사 자료는 에스케이텔레콤에서 제작하였습니다.

· 데이터 구성

● 데이터 셋 항목: 데이터는 TARGET, SENTENCE, ANSWER로 구성되어 있습니다.

- TARGET : 2개 이상의 의미를 가진 대상 단어
- SENTENCE : 대상 단어(TARGET)를 포함한 2개의 문장(SENTENCE1, SENTENCE2)을 직접 생성하거나 혹은 사전의 예시문, 신문 기사 등에서 추출
- ANSWER: True/False로 구성. SENTENCE 안에 포함된 대상 단어(TARGET)가 완벽하게 동일한 의미로 사용된 경우는 True로, 문장 안에서 완전히 다른 의미로 사용된 경우 False로 태깅함.

● 데이터 셋 생성 주요 조건

- 대상 단어인 TARGET은 동일한 형태이지만 두 개 이상의 의미로 사용이 되는 것으로 선택합니다.
 - 주로 동형이의어와 같이 형태가 같지만 의미가 완전히 다른 경우들로 구성합니다.
 - 표준국어대사전 또는 우리말샘 등재어라면, 줄임말이나 구어체 등도 대상 단어로 선택할 수 있습니다.
- SENTENCE 1과 SENTENCE 2는 모두 문장 내에서의 컨텍스트만으로 TARGET의 의미가 무엇인지 명확히 파악 가능하도록 추출하거나 생성하여야 합니다.

Target: 김

SENTENCE 1: 밥을 **김**에 싸서 먹다.

SENTENCE 2: 안경알에 **김**이 끼다.

ANSWER: **False**

Target: 수면

SENTENCE 1: 잔잔한 **수면**에 얼굴을 비추다.

SENTENCE 2: 적당한 강바람에 배가 **수면** 위를 부드럽게 미끄러져 갔다.

ANSWER: **True**

· 자료 규모 및 평가 지표

구분	훈련(Train)	개발(Dev)	시험(Test)	평가 지표
내용	7748셋	1166셋	1246셋	Accuracy

· 파일 형식: TSV (UTF-8)

· 예시

ID	TARGET	SENTENCE 1	SENTENCE 2	ANSWER
1	발	[발]을 헛디더 발목이 빠었다.	경찰은 달아나는 범인을 향해 공포탄을 한 [발] 쏘았다.	False
2	김	그 마을에는 계단식 논 한가운데 [김] 건조실이 있었다.	희선이가 [김]이 나는 떡국을 들고 방으로 들어왔다.	False
3	수면	밀물과 썰물의 현상에 따라 바다의 [수면]은 색과 높이가 변한다고 한다.	잔잔한 [수면]에 얼굴을 비추다.	True
4	연초	[연초]부터 모든 공공요금이 크게 인상되었다.	새해 [연초]에 토정비결을 보았다.	True
5	이르	롯데폰기의 열기는 2일 오후 한국 가수들이 출연한 케이팝 콘서트에서 절정에 [이르]렀다.	이순신은 장수들을 불러 모아 “병법에 [이르]기를 죽으려 하면 살고 살려고 하면 죽는다”고 말했다.	FALSE
6	그리	스스로를 [그리]는 자화상 역시 자신을 알고자 하는 욕망이 담겼다.	"신부님, 성인들은 늘 천국을 [그리]면서 사셨다는데 정말 안 가고 싶으세요?"	FALSE
7	맞	특히 올해 창단 8주년을 [맞]은 봉사단은 나눔실천 유공기업으로 선정되기도 했다.	석양이 질 무렵 위기를 [맞]았다.	TRUE
8	강하	이에 비주류가 [강하]게 반발해, 당 내분이 격화하고 있다.	우리가 그를 더 [강하]게 만들어줬다.	TRUE