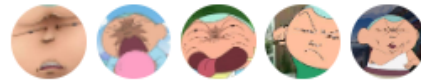


NLP 8조 MIML Solution 발표

Data Annotation



T3180 임동진
T3201 정재윤
T3205 조설아
T3238 허치영
T3246 이보림

Contents

Overview

제작 과정

Problems

Entity Type & Relation Map

Results

후기

Overview

- Topic : 반려동물
- 출처 : 한국어 위키피디아
- 텍스트 파일 수 : 81
- 총 line 수 : 2193
- 최종 sentence 수 : 1755



몽구스
빈투롱
아프리카사향고양이
오리너구리
오리너구리과
장수거북
제넷속

너희도 반려동물이니...?

제작 과정










Label Studio



Entity Types [docs](#)

Here you define the categories to highlight in text. An entity can be any concept, for example: protein name, risk, organization, paragraph highlight, customer feedback, etc.

To automatically annotate and disambiguate entities, use [Dictionaries](#) or/and [Machine Learning](#).

	<input type="text" value="sbj-anm-habitat"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="obj-loc-habitat"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="sbj-anm-alternate_name"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="obj-poh-alternate_name"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="obj-org-alternate_name"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="obj-anm-alternate_name"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>
	<input type="text" value="sbj-anm-physical"/>	<input type="text" value="Optional description"/>	<input type="button" value="Delete"/>	<input type="button" value="Save"/>





사모예드견은 순하고 **친근**하고 솔직하며, 기민하고 매우 발랄하다.



Entities

total **2** not normalized **2**

[Group/filter entities](#) ▼

 sbj-anm-habit	1	◀
 1 ⚡ 0 (0.00%) ↔ 1		
 사모예드견	1	▼
 obj-poh-habit	1	◀
 1 ⚡ 0 (0.00%) ↔ 1		
 친근	1	▼

Relations

total **1**

 친근 ↔  사모예드견	✕
--	---

Label Studio Projects / Project / Labeling

#728 ID idj7183 #280 created, 2 days ago 1/1

text

저먼 셰퍼드는 제1차 세계 대전이 일어날 때부터 독일 전역에서 군용견으로 훈련을 받아 크게 활약했었다.

SUB_ANM 1 SUB_ORG 2 OBJ_LOC 3 OBJ_ANM 4 OBJ_POH 5

OBJ_ORG 6 OBJ_NOH 7 OBJ_PER 8 OBJ_DAT 9

anm:역할

Update

No Region selected

Regions 2 Labels

Sorted by Date

1 SUB_ANM 저먼 셰... 2 OBJ_ORG 군용견

Relations (1)

1 HTML → 2 HTML

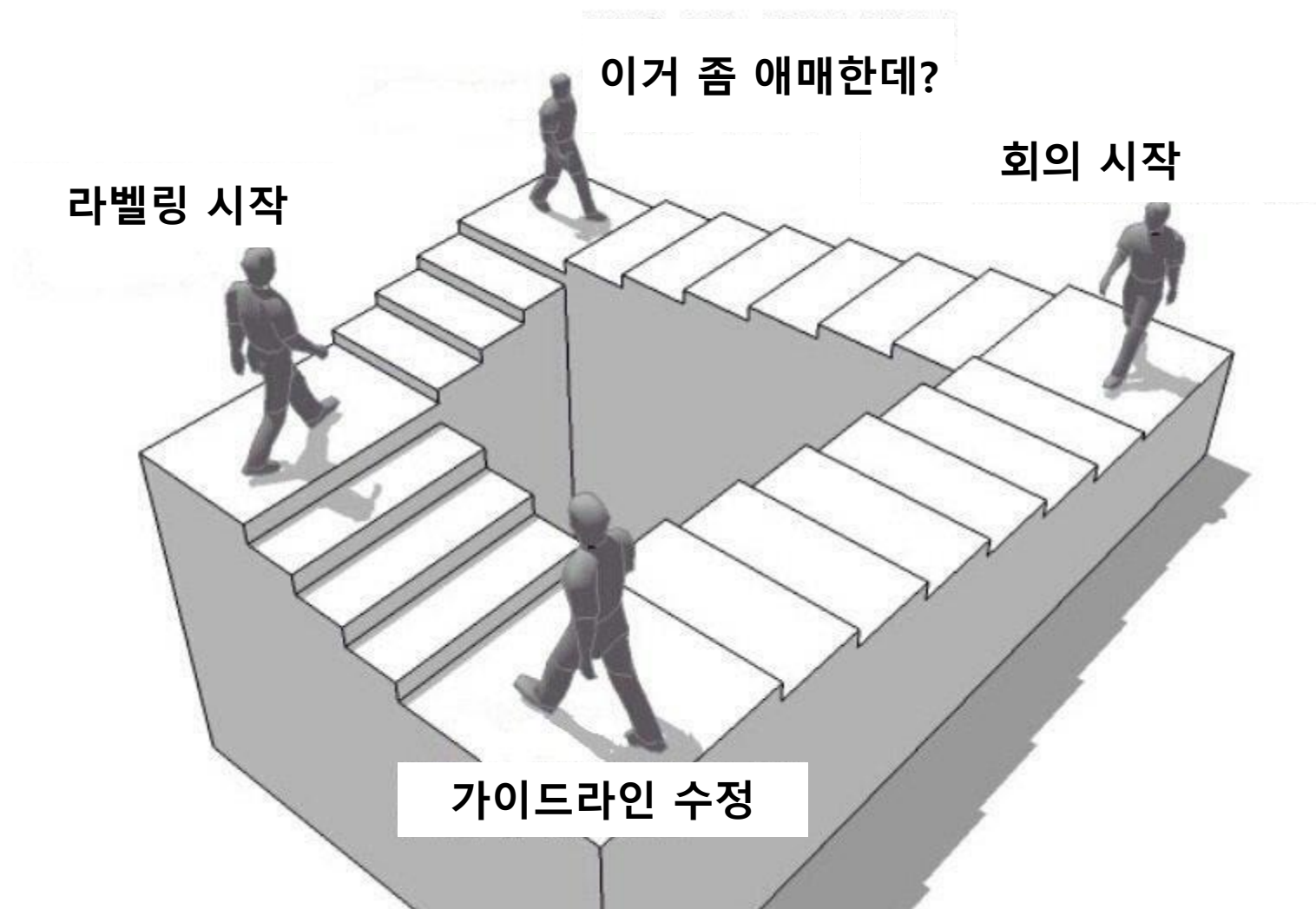
LABELS

anm:역할 ×

Tagtog이 relation 설정이 너무 까다로워서 label studio를 썼다.
(말로하기)

sentence	sentence_with_entity	subject_entity	object_entity	class
아프리카들고양이는 본래 들고양이의 아종으로 분류된다	<obj:아프리카들고양이>는 본래 <sbj:들고양이의 아종>으로 분류된다	{'start': 13, 'end': 20, 'text': '들고양이의 아종', 'type': 'org'}	{'start': 0, 'end': 7, 'text': '아프리카들고양이', 'type': 'org'}	no_relation
아프리카들고양이는 고양이속에 속한 소형 고양이과 동	아프리카들고양이는 <sbj:고양이속>에 속한 <obj:소형 고양이과> 동	{'start': 10, 'end': 13, 'text': '고양이속', 'type': 'org'}	{'start': 19, 'end': 25, 'text': '소형 고양이과', 'type': 'org'}	org:sub_group
아프리카들고양이는 아프리카·서아시아·중앙아시아·인	<sbj:아프리카들고양이>는 아프리카·서아시아·중앙아시아·인도·<obj:	{'start': 0, 'end': 7, 'text': '아프리카들고양이', 'type': 'anm'}	{'start': 29, 'end': 30, 'text': '중국', 'type': 'loc'}	anm:habitat
대한민국에 서식하는 청개구리는 청개구리 및 수원청개	<obj:대한민국>에 서식하는 청개구리는 청개구리 및 <sbj:수원청개구	{'start': 24, 'end': 29, 'text': '수원청개구리', 'type': 'anm'}	{'start': 0, 'end': 3, 'text': '대한민국', 'type': 'loc'}	anm:habitat
청개구리는 주로 평지와 저지대에 서식하며, 번식기 이	<sbj:청개구리>는 주로 평지와 <obj:저지대>에 서식하며, 번식기 이	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 13, 'end': 15, 'text': '저지대', 'type': 'loc'}	anm:alternate_name anm:physical
청개구리는 등쪽이 초록색을 띄고 있지만 항상 그런 것	<sbj:청개구리>는 <obj:등쪽이 초록색>을 띄고 있지만 항상 그런 것	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 6, 'end': 12, 'text': '등쪽이 초록색', 'type': 'poh'}	anm:habit
청개구리는 비슷한 소리가 연달아 들리면 뒷소리를 무	<sbj:청개구리>는 비슷한 소리가 연달아 들리면 뒷소리를 무시하고	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 49, 'end': 54, 'text': '선행음 효과', 'type': 'poh'}	anm:enemy
청개구리는 앞다리의 발가락 기부에 흔적적인 물갈퀴>가	<sbj:청개구리>는 앞다리의 발가락 기부에 흔적적인 <obj:물갈퀴>가	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 24, 'end': 26, 'text': '물갈퀴', 'type': 'poh'}	anm:role
청개구리는 죽은 나무 밑에서 겨울잠을 자고 번식기는	<sbj:청개구리>는 죽은 나무 밑에서 겨울잠을 자고 번식기는 <obj:5-	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 29, 'end': 33, 'text': '5-6월경', 'type': 'noh'}	anm:origin
수컷은 인도 부근에 커다란 울음주머니가 있고 암컷은	<sbj:수컷>은 인도 부근에 커다란 <obj:울음주머니>가 있고 암컷은	{'start': 0, 'end': 1, 'text': '수컷', 'type': 'anm'}	{'start': 15, 'end': 19, 'text': '울음주머니', 'type': 'poh'}	anm:number_of_members
청개구리는 환경에 따라 현저하게 색이 변하기도 한다	<sbj:청개구리>는 환경에 따라 현저하게 <obj:색이 변하기>도 한다.	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 18, 'end': 23, 'text': '색이 변하기', 'type': 'poh'}	anm:physical
청개구리는 낮에는 숲 속에서 조용히 있으나, 밤이 되	<sbj:청개구리>는 낮에는 숲 속에서 조용히 있으나, 밤이 되면 논가로	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 58, 'end': 59, 'text': '수컷', 'type': 'anm'}	no_relation
청개구리의 알은 논이나 못 등 고인 물의 물풀 같은 곳	<sbj:청개구리>의 알은 논이나 못 등 고인 물의 물풀 같은 곳에 붙어	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 67, 'end': 74, 'text': '1-10개의 알', 'type': 'noh'}	no_relation
청개구리(Japanese tree frog)는 청개구리과의 개구리	<obj:청개구리(Japanese tree frog)>는 <sbj:청개구리과>의 개구리	{'start': 26, 'end': 30, 'text': '청개구리과', 'type': 'org'}	{'start': 0, 'end': 23, 'text': '청개구리(Japanese tree frog)', 'type': 'a	org:sub_group
청개구리는 나무나 풀, 숲에 있을 때는 녹색을 띄지만,	<sbj:청개구리>는 나무나 풀, 숲에 있을 때는 녹색을 띄지만, 땅에 있	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 51, 'end': 56, 'text': '흑색의 무늬', 'type': 'poh'}	anm:physical
청개구리의 먹이는 주로 애벌레, 거미 등 절지동물을 잡	<sbj:청개구리>의 먹이는 주로 애벌레, 거미 등 <obj:절지동물>을 잡	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 23, 'end': 26, 'text': '절지동물', 'type': 'anm'}	anm:enemy
청개구리의 등면은 녹색이나 황록색 바탕에 진한 녹색	<sbj:청개구리>의 등면은 녹색이나 황록색 바탕에 진한 녹색 또는 <	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 32, 'end': 43, 'text': '흑갈색의 불규칙한 무늬', 'type': 'poh'}	anm:physical
청개구리의 몸길이는 25-40mm 정도이다.	<sbj:청개구리>의 몸길이는 <obj:25-40mm> 정도이다.	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 11, 'end': 17, 'text': '25-40mm', 'type': 'noh'}	anm:physical
청개구리는 다른 양서류와 달리 앞뒤 발가락에 빨판이	<sbj:청개구리>는 다른 양서류와 달리 앞뒤 <obj:발가락에 빨판>이	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 20, 'end': 26, 'text': '발가락에 빨판', 'type': 'poh'}	anm:physical
청개구리가 큰소리로 짹 찾으면서 천적을 피할 수 있는	<sbj:청개구리>가 큰소리로 짹 찾으면서 천적을 피할 수 있는 비결은	{'start': 0, 'end': 3, 'text': '청개구리', 'type': 'anm'}	{'start': 50, 'end': 53, 'text': '울음소리', 'type': 'poh'}	anm:habit
2020년 10월 발표된 사이언스 논문에 의하면 진돗개의	2020년 10월 발표된 사이언스 논문에 의하면 <sbj:진돗개>의 유전 정	{'start': 27, 'end': 29, 'text': '진돗개', 'type': 'anm'}	{'start': 45, 'end': 51, 'text': '늑기니 고산개', 'type': 'anm'}	no_relation
진돗개(珍島개)는 개 품종 중 하나로, 한국의 대표적	<obj:진돗개(珍島개)>는 개 품종 중 하나로, 한국의 대표적인 <sbj:	{'start': 33, 'end': 36, 'text': '개 품종', 'type': 'org'}	{'start': 0, 'end': 9, 'text': '진돗개(珍島개)', 'type': 'anm'}	org:sub_group
진돗개는 어미에 대한 모성애가 매우 강하다.	<sbj:진돗개>는 어미에 대한 <obj:모성애>가 매우 강하다.	{'start': 0, 'end': 2, 'text': '진돗개', 'type': 'anm'}	{'start': 12, 'end': 14, 'text': '모성애', 'type': 'poh'}	anm:habit
일반적으로 수컷은 암컷보다 머리가 크며 암컷은 좀 더	일반적으로 <sbj:수컷>은 암컷보다 <obj:머리가 크며> 암컷은 좀 더	{'start': 6, 'end': 7, 'text': '수컷', 'type': 'anm'}	{'start': 15, 'end': 20, 'text': '머리가 크며', 'type': 'poh'}	anm:physical

좀 더 편안한 Labeling!!



Problems

- 동물: 서식지 (anm: habitat)
- 동물: 대체표현 (anm: alternate_name)
- 동물: 신체적 특징 (anm: physical)
- 동물: 질병 (anm: disease)
- 동물: 비신체적 특징 (anm: habit)
- 동물: 사냥감 (anm: enemy)
- 동물: 역할 (anm: role)
- 동물: 유래 (anm: origin)
- 동물: 개체수 (anm: number_of_members)
- 집단: 상위집단 (org: upper_group)
- 집단: 하위집단 (org: sub_group)
- 관계 없음 (no_relation)

예시는 읽지말자

“종”의 개념을 사전지식으로 봐야 될까?

- **고슴도치**는 몸길이는 약 20cm 안팎으로 **대형종**이다.
- **도롱뇽**은 **서울시 보호 야생 생물 대상종**이다.
- **달마시안**은 크로아티아의 달마티아가 원산지인 개 **품종**이다.

“품종”, “개체”를 동물(ANM)이라고 봐도 될까?

- 이 **품종**은 주인과 유대감이 빨리 형성돼 혼자 시간을 보낼 수 있도록 훈련 시키지 않는다면, **분리 불안**에 시달릴 수 있다.
- 이 **개체**는 대체로 육식을 하며, 작은 포유류(저빌, 산토끼류, 설치류)와 **새** 등을 좋아한다.

“질병”을 신체적 특징으로 분류해도 될까?

- **풍산개**의 주요 질병으로는 요로결석, **복막염**, 폐암, 심장사상충, 백내장, 탈모증 등이 있다.



이가람_멘토 오후 8:56

스레드에 답글을 남김: 멘토님 저희 질문이 하나 더 나왔는데, 사전지식 허용 범위에 관한 질문이 있습니다....

[KLUE-RE 관계 추출 태스크 가이드라인]에 보시면 2-3) 외부 지식의 활용 항목이 있는데요.

본 태스크는 반드시 주어진 문장에 나타나는 정보만을 활용하여 Entity 쌍의 관계를 유추해내야 합니다. 따라서 문장에서 얻을 수 없는 정보, 외부 지식을 활용해서 관계를 유추하면 안 됩니다.

우선 가이드라인에서 제시한 것과 같이 외부 지식을 활용하지 않고 문장에서 얻을 수 있는 정보로만 판단이 이루어져야 하는데요

제가 봤을 때는 '과'라는 단어 자체(종속과목강관계)가 뭔지 모르더라도 반려동물 데이터에서 몇가지 예시를 보았을 때

속한다/중 하나이다/떨린 등의 문장 내 표현과 항상 같이 쓰이는 것 같아 문장 내의 정보만으로 '하위집단'이라는 것이 판단이 가능하지 않을까 생각합니다.

따라서 제 생각은 외부 사전지식을 활용하면 안되고 문장 내의 정보로만 파악을 해야하는데

- '종속과목강관계'가 언급되어있을 때 주로 속한다/중 하나이다/떨린 등의 표현과 같이 쓰이므로 이런 표현과 함께 쓰이면 '하위집단'에 해당된다고 생각되구요
- 다만, 속한다/중 하나이다/떨린 등의 표현이 없이 '종속과목강관계'만 딱 언급된다면 '관계 없음'이라고 판단하는게 맞지 않을까 생각합니다.
 - 예시
 - 고양이(sub)는 고양이과(obj)에 속한다 : 하위집단
 - 고양이(sub)는 고양이과(obj)이다 : 관계없음

'집단: 상위_집단, 집단: 하위_집단' relation → '집단: 하위_집단}' 하나로 통일하 것이

깔끔할 것 같습니다.

- subject 와 object를 통해서 상/하위를 지정해줄 수 있기 때문에 두 관계를 모두 설정하는 것은 불필요할 수 있습니다.



조원익_조교 오후 2:11

스레드에 답글을 남김: 안녕하세요 NLP 8조 임동진입니다. 애완동물에 관한 데이터 제작을 하고 있는데 대명사에 대한 질문이 있습니다...

안녕하세요. 비슷한 이슈가 생겼던 '자연과학' 주제에서는 이러한 점을 고려하여 '개체'라는 entity를 정의하면 어떠한지에 대해 이야기를 드린 바 있습니다. 또 '반려동물'이라는 topic 하에서 나오는 개체는 대부분 동물이라는 점을 고려하여, 동물로 간주해도 저는 괜찮다고 생각합니다. 해당 가이드라인에서의 entity 체계가 일반 topic까지 모두 적용될 필요는 없습니다. 또한 말씀하신 대로, 고양이 라는 일반적인 생물종으로 문장을 수정하셔도 괜찮습니다 (article이 고양이에 관한 것이라면) relation이 도출되기 힘들어보인다면 원문을 고수하실 필요는 없습니다 😊



1



- 동물: 서식지 (ANM: habitat)
- 동물: 대체표현 (ANM: alternate_name)
- 동물: 신체적 특징 (ANM: physical)
- ~~동물: 질병 (anm: disease)~~
- 동물: 비신체적 특징 (ANM: habit)
- 동물: 사냥감 (ANM: enemy)
- 동물: 역할 (ANM: role)
- 동물: 유래 (ANM: origin)
- 동물: 개체수 (ANM: number_of_members)
- ~~집단: 상위집단 (ORG: upper_group)~~
- 집단: 하위집단 (ORG: sub_group)
- 관계 없음 (no_relation)

	A	B	C	D	E	F	G	H	I	J	K
1	ser	sentence_with_entity	subject_entity	object_entity	치영	동진	보림	재윤	설아	일치	Final
2	사모0	<sbj:사모예드건>은 <obj:시베리아 원산>으로 썰매용·가정용·반려용으로 많이 기른다.	{'start': 0, 'end': 4, 'text': '사모예드건'}	{'start': 7, 'end': 13, 'text': '시베리아 원산'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
3	그러나	그러나 1960년대 말 경북대 교수들에 의해 <obj:30여 마리>의 <sbj:삽살개>가 수집, 보존되	{'start': 33, 'end': 35, 'text': '30여 마리'}	{'start': 25, 'end': 30, 'text': '삽살개'}	anm:number_of_members	anm:number_of_members	anm:number_of_members	anm:number_of_members	anm:number_of_members	TRUE	anm:number_of_members
4	동네	동네마다 흔하던 <sbj:삽살개>는 <obj:일제 강점기>인 1940년 이후 일본이 개를 전쟁에 필	{'start': 9, 'end': 11, 'text': '일제 강점기'}	{'start': 14, 'end': 19, 'text': '1940년 이후'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
5	삽살개	<sbj:삽살개>는 <obj:한반도>의 동남부 지역에 널리 서식하던 대한민국의 토종개이다.	{'start': 0, 'end': 2, 'text': '한반도'}	{'start': 5, 'end': 7, 'text': '동남부 지역'}	anm:habitat	anm:habitat	anm:habitat	anm:habitat	anm:habitat	TRUE	anm:habitat
6	삽살개	<sbj:삽살개>는 <obj:신라시대>에는 주로 귀족사회에서 길러져 오다가 통일신라가 망하면서	{'start': 0, 'end': 2, 'text': '신라시대'}	{'start': 5, 'end': 8, 'text': '귀족사회'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
7	귀신	<obj:귀신과 액운을 쫓는 개>라는 뜻을 지닌 <sbj:삽살개>는 이름 자체도 순수한 한국어로	{'start': 23, 'end': 25, 'text': '귀신과 액운을 쫓는 개'}	{'start': 0, 'end': 13, 'text': '삽살개'}	anm:alternate_name	anm:alternate_name	anm:origin	anm:origin	anm:origin	FALSE	anm:alternate_name
8	그러나	그러나 2차 세계대전으로 인해 <sbj:프랑스>의 <obj:자생 블루 고양이 군락>은 자취를 감추	{'start': 17, 'end': 19, 'text': '프랑스'}	{'start': 22, 'end': 33, 'text': '자생 블루 고양이 군락'}	org:sub_group	no_relation	anm:habitat	anm:origin	org:sub_group	FALSE	org:sub_group
9	미국	미국의 경우, <obj:1970년대> <sbj:샤르트뢰>를 들여오게 되었다.	{'start': 15, 'end': 18, 'text': '1970년대'}	{'start': 8, 'end': 13, 'text': '샤르트뢰'}	anm:origin	anm:origin	anm:origin	anm:origin	no_relation	FALSE	anm:origin
10	프랑스	프랑스 카르투지오 수도회 수사들이 <obj:샤르투르즈>라는 리큐어를 개발하였는데 이것이	{'start': 45, 'end': 50, 'text': '샤르투르즈'}	{'start': 19, 'end': 25, 'text': '리큐어'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
11	샤르트	<sbj:샤르트뢰>는 20세기 두 차례의 세계대전을 거치면서 멸종 위기에 처하게 되었으나,	{'start': 0, 'end': 3, 'text': '샤르트뢰'}	{'start': 45, 'end': 47, 'text': '20세기'}	no_relation	no_relation	anm:origin	anm:origin	anm:origin	FALSE	anm:origin
12	이런	이런 노력으로 1928년에 <sbj:샤르트뢰>는 다시 유럽 <obj:고양이쇼>에 등장하기 시작하였	{'start': 15, 'end': 18, 'text': '1928년'}	{'start': 27, 'end': 30, 'text': '고양이쇼'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
13	또한	또한 <sbj:스핑크스 고양이>는 솜털이 매우 짧아 체온이 사람의 손에 그대로 전달되어 다른	{'start': 3, 'end': 10, 'text': '스핑크스 고양이'}	{'start': 58, 'end': 61, 'text': '사람의 손'}	anm:physical	anm:physical	no_relation	anm:physical	anm:physical	FALSE	no_relation
14	기록	기록을 보면 1902년 <obj:멕시코>를 시점으로 여러 군데에서 <sbj:털이 없는 고양이>가 발	{'start': 31, 'end': 39, 'text': '멕시코'}	{'start': 13, 'end': 15, 'text': '털이 없는 고양이'}	anm:origin	anm:origin	anm:origin	anm:origin	anm:origin	TRUE	anm:origin
15	이 중	이 중을 어려서부터 키우는 사람은 자신의 <sbj:스핑크스 고양이>에게 맞는 사료를 찾아내	{'start': 23, 'end': 30, 'text': '스핑크스 고양이'}	{'start': 96, 'end': 97, 'text': '사료'}	no_relation	no_relation	no_relation	no_relation	no_relation	TRUE	no_relation
16	보통	: 보통 <sbj:스핑크스 고양이>는 단색으로서 깨끗한 분홍색상이 소위 <obj:올핑크>라 하여	{'start': 3, 'end': 10, 'text': '스핑크스 고양이'}	{'start': 32, 'end': 34, 'text': '올핑크'}	anm:alternate_name	anm:physical	anm:physical	anm:physical	anm:physical	FALSE	anm:physical
17	또한	또한 <sbj:스핑크스 고양이>는 다른 고양이들에 비해 특별한 이유가 없는 한 주인에게	{'start': 3, 'end': 10, 'text': '스핑크스 고양이'}	{'start': 43, 'end': 45, 'text': '주인'}	no_relation	anm:habit	no_relation	anm:habit	no_relation	FALSE	no_relation
18	오늘날	오늘날 고양이 혈통서를 보면 <obj:에피데미스>라는 이름이 기록된 <sbj:스핑크스>가 많은	{'start': 34, 'end': 37, 'text': '에피데미스'}	{'start': 16, 'end': 22, 'text': '스핑크스'}	anm:alternate_name	anm:alternate_name	anm:origin	anm:origin	anm:origin	FALSE	anm:alternate_name
19	스핑크	<sbj:스핑크스>는 모습이 <obj:스핑크스>와 모습이 비슷하여 이름이 스프링크스 이다.	{'start': 0, 'end': 3, 'text': '스핑크스'}	{'start': 10, 'end': 13, 'text': '스핑크스'}	no_relation	anm:origin	no_relation	anm:origin	anm:origin	FALSE	anm:origin
20	그러나	그러나 다른 품종의 <sbj:고양이>와는 비교할 수 없을 정도로 눈에 보이는 <obj:털>이 없기	{'start': 11, 'end': 13, 'text': '고양이'}	{'start': 37, 'end': 37, 'text': '털'}	no_relation	no_relation	no_relation	anm:physical	anm:physical	FALSE	anm:physical
21	날찍	날찍한 가슴과 배 부분은 오크통에 가까운 형태로 둥글고 볼록해 보이지만 비만으로 보일 정	{'start': 65, 'end': 69, 'text': '가슴과 배'}	{'start': 75, 'end': 79, 'text': '오크통'}	no_relation	no_relation	anm:physical	no_relation	no_relation	FALSE	no_relation
22	짧은	짧은 털을 가진 <sbj:고양이>들은 2차례 자연적 <obj:돌연변이>를 일으켜 생겨났다.	{'start': 9, 'end': 11, 'text': '고양이'}	{'start': 23, 'end': 26, 'text': '돌연변이'}	anm:physical	anm:physical	anm:origin	anm:physical	anm:origin	FALSE	anm:physical
23	때문	때문에 <sbj:스핑크스 고양이>는 <obj:실내>에서 키우는 것이 좋고, 보온에 신경을 써야 한	{'start': 4, 'end': 11, 'text': '스핑크스 고양이'}	{'start': 14, 'end': 15, 'text': '실내'}	no_relation	no_relation	anm:habitat	no_relation	no_relation	FALSE	no_relation
24	특히	특히 <sbj:스핑크스>는 이마에 주름이 잡혀있어 뭔가 걱정이 많은 듯한 특유의 <obj:노인네	{'start': 3, 'end': 6, 'text': '스핑크스'}	{'start': 39, 'end': 47, 'text': '노인네'}	anm:habit	anm:habit	anm:habit	anm:physical	anm:habit	FALSE	anm:habit
25	스핑크	<sbj:스핑크스 고양이>를 보면 <obj:피부>를 만져보고 싶은 생각이 든다는 사람이 많다.	{'start': 0, 'end': 7, 'text': '스핑크스 고양이'}	{'start': 13, 'end': 14, 'text': '피부'}	no_relation	anm:habit	no_relation	anm:physical	no_relation	FALSE	no_relation
26	스핑크	<obj:스핑크스>는 <sbj:단모종> (털이 짧은) 고양이의 품종 중 하나로 1998년 CFA에 잡종부	{'start': 6, 'end': 8, 'text': '단모종'}	{'start': 0, 'end': 3, 'text': 'CFA'}	org:sub_group	org:sub_group	org:sub_group	org:sub_group	org:sub_group	TRUE	org:sub_group
27	첫번째	첫번째 돌연변이 발생은 미국 미네소타 주 와디너의 어느 농장에서 떠돌이 고양이 <obj:제	{'start': 68, 'end': 73, 'text': '미네소타 주 와디너'}	{'start': 44, 'end': 48, 'text': '제'}	anm:origin	no_relation	anm:origin	no_relation	anm:origin	FALSE	anm:origin
28	스핑크	<sbj:스핑크스 고양이>는 발정기의 <obj:쿨링 목소리>도 그리 심한 편이 아니고, 여러 마리	{'start': 0, 'end': 7, 'text': '스핑크스 고양이'}	{'start': 15, 'end': 20, 'text': '쿨링 목소리'}	anm:habit	anm:habit	anm:habit	anm:physical	anm:habit	FALSE	anm:habit
29	스핑크	<sbj:스핑크스 고양이>는 <obj:장점>이 많은 고양이인 반면에 키우면서 신경써야 할 일도	{'start': 0, 'end': 7, 'text': '스핑크스 고양이'}	{'start': 10, 'end': 11, 'text': '장점'}	no_relation	anm:habit	no_relation	no_relation	no_relation	FALSE	no_relation
30	1978	1978년에는 털이 없는 수컷 새끼 고양이 1마리와 암컷 새끼 고양이 2마리가 토론토 거리	{'start': 82, 'end': 85, 'text': '1978년'}	{'start': 65, 'end': 67, 'text': '토론토 거리'}	no_relation	anm:origin	no_relation	anm:origin	no_relation	FALSE	no_relation

- 동물: 서식지 (ANM: habitat)
- 동물: 대체표현 (ANM: alternate_name)
- 동물: 신체적 특징 (ANM: physical)
- 동물: 비신체적 특징 (ANM: habit)
- 동물: 사냥감 (ANM: enemy)
- 동물: 역할 (ANM: role)
- 동물: 유래 (ANM: origin)
- 동물: 개체수 (ANM: number_of_members)
- 집단: 하위집단 (ORG: sub_group)
- 관계 없음 (no_relation)

physical? habit?

카멜레온의 특징은 **피부 색**을 바꿀 수 있다는 것인데, 변하는 색은 도마뱀의 물리적이고 심리적인 표현이며, 보통 알려진 것처럼 주변 환경에 맞춘다.

physical? habit?

이 **품종**은 주인과 유대감이 빨리 형성돼 혼자 시간을 보낼 수 있도록 훈련 시키지 않는다면, **분리 불안**에 시달릴 수 있다.

no_relation? origin?

여왕의 일생 동안 **포메라니안** 품종의 크기는 50%까지 줄었다.



그렇게 6시간 후...

physical? habit?

카멜레온의 특징은 **피부 색**을 바꿀 수 있다는 것인데, 변하는 색은 도마뱀의 물리적이고 심리적인 표현이며, 보통 알려진 것처럼 주변 환경에 맞춘다.

physical? habit?

이 **품종**은 주인과 유대감이 빨리 형성돼 혼자 시간을 보낼 수 있도록 훈련 시키지 않는다면, **분리 불안**에 시달릴 수 있다.

no_relation? origin?

여왕의 일생 동안 **포메라니안** 품종의 크기는 50%까지 줄었다.

origin

- 역사, 설화, 속담, 이야기, 유래 등의 과거 기록에 연관된 인물, 사건, 지역 등
- 문화와 연관된 인물, 사건, 지역 등

physical

- 해부학적으로 측정 및 해석 가능한 표현
- 신체 기관에 의해 발생하는 현상 및 특징 (체온, 시각 등)

habit

- 동물의 습성
- 명확한 해부학적 정의가 불가능한 경우 (산란 장소, 수명 등)
- 성격, 표정 등 주관에 의해 판단하게 되는 특징
- 습성, 행동을 설명하기 위한 신체 부위
- 동물-사람의 관계
- 부재의 표현은 해당하지 않음 (없다, 거의 ~않는다 등)

사전지식

- 동물 분류 체계(종속과목강문계)는 포함 및 종속에 관한 표현이 없는 경우에는 “관계 없음” 으로 해석

Entity Type & Relation Map

- **ANM**
 - 독립적으로 서술된 동물
 - 동물의 부산물(알)은 해당하지 않음
- **ORG**
 - 개체수 관련 관계, 종속 관계인 집단 및 개체
 - 동물 분류 체계 명칭, 동물의 품종 및 등급
 - 서식지나 경유지로 사용되지 않는 하나의 지명 혹은 단체
- **POH**
 - 위의 엔티티 타입들에 속하지 않는 엔티티 타입
 - 신체 부위, 식물, 성분, 질병 및 역할에 관련된 명사
- **NOH**
 - 동물의 마리 수, 개수, 기온, 길이 혹은 무게의 단위
- **PER**
 - 사람의 이름, 직위 혹은 두 가지 모두 사용되는 경우
- **DAT**
 - 기간이나 특정 시점에 관련된 엔티티
- **LOC**
 - 지역, 지명이라 하더라도 다른 Entity type으로 명확히 구분 가능한 경우를 제외하면 LOC로 지정

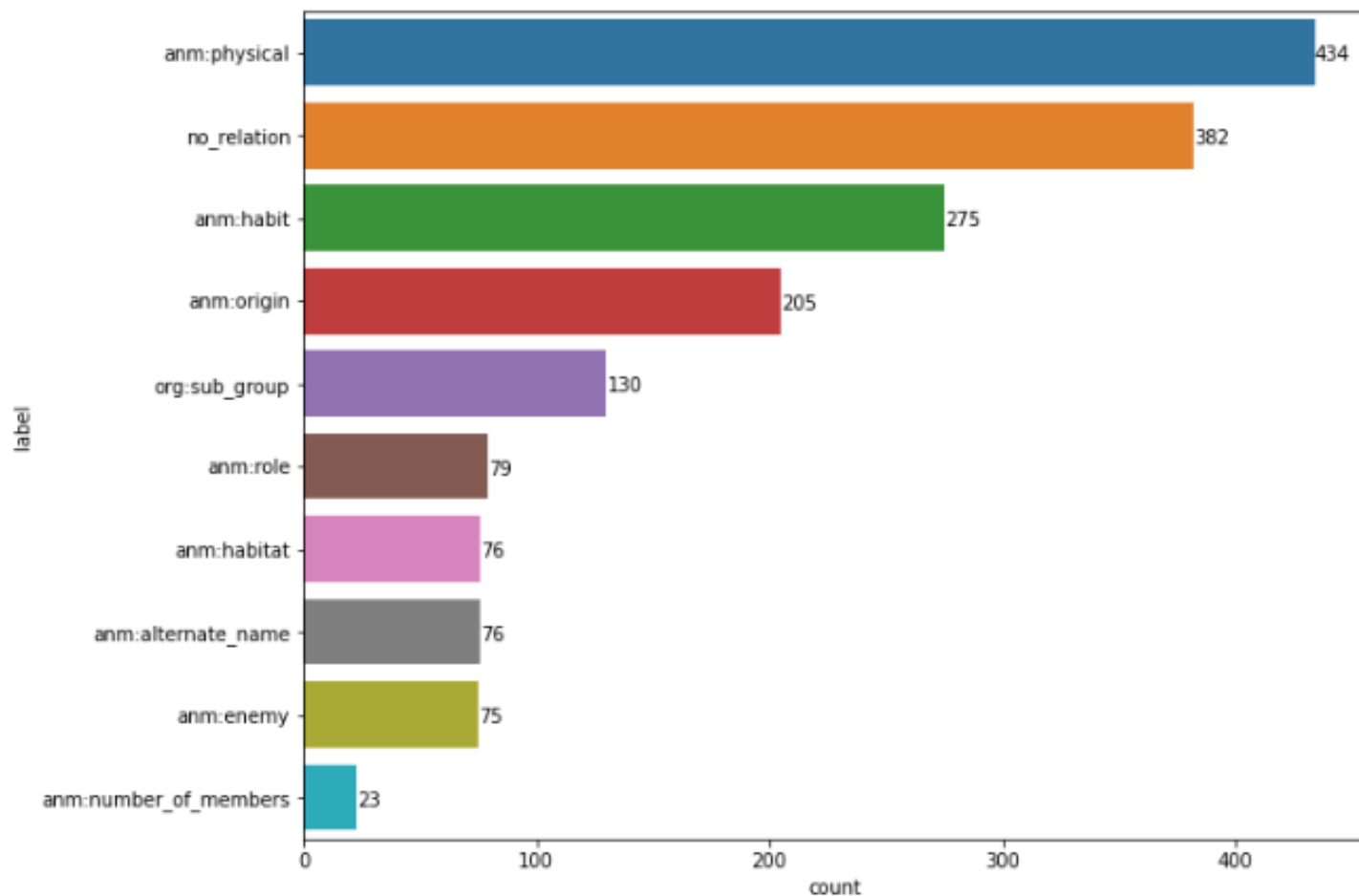
- 동물: 서식지 (anm: habitat)
 - (ANM, LOC)
- 동물: 대체표현 (anm: alternate_name)
 - (ANM, ANM/POH/ORG)
- 동물: 신체적 특징 (anm: physical)
 - (ANM, POH/NOH)
- 동물: 비신체적 특징 (anm: habit)
 - (ANM, POH/NOH/DAT)
- 동물: 사냥감 (anm: enemy)
 - (ANM, ANM/POH)
- 동물: 역할 (anm: role)
 - (ANM, ORG/POH)
- 동물: 유래 (anm: origin)
 - (ANM, ANM/ORG/POH/PER/LOC)
- 동물: 개체수 (anm: number_of_members)
 - (ANM, NOH)
- 집단: 하위집단 (org: sub_group)
 - (ORG, ANM/ORG)
- 관계 없음 (no_relation)
 - (ANM/ORG, *)

Results

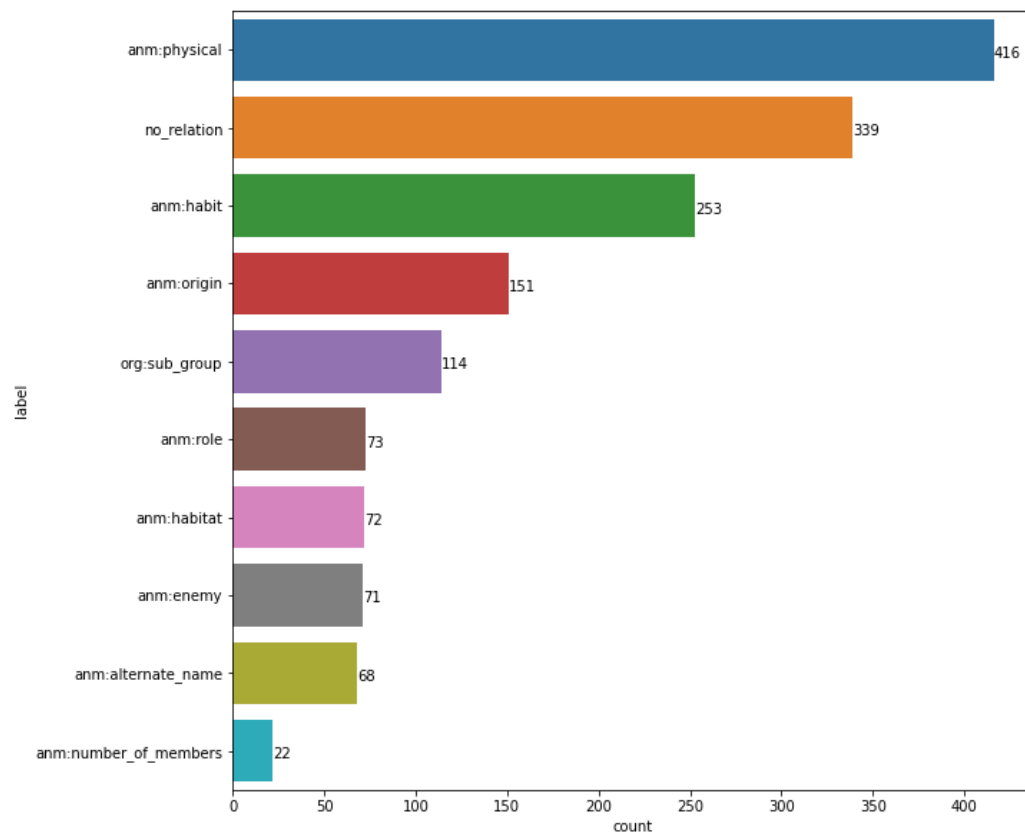
- **Pilot Tagging** : 0.72
- **Final Tagging** : 0.79

```
!python calculate_iaa.py
```

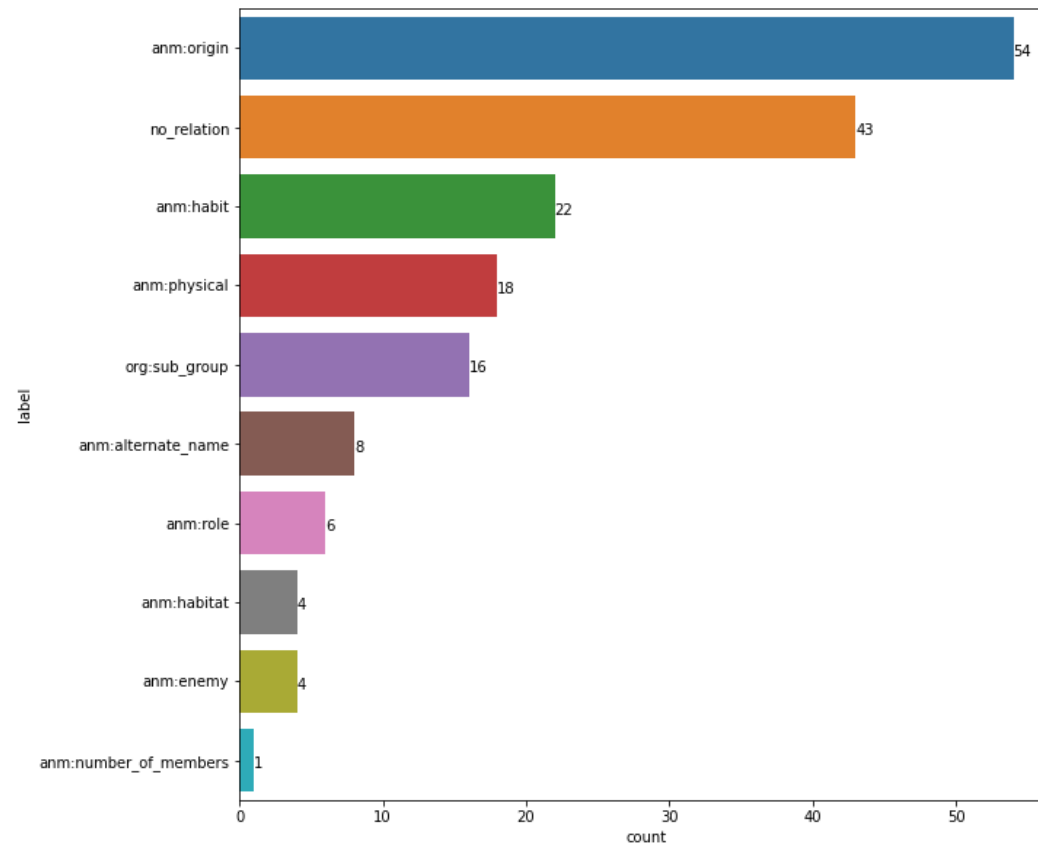
```
#raters = 5 , #subjects = 1601 , #categories = 9  
PA = 0.8309181761399111  
PE = 0.19568345986132185  
Fleiss' Kappa = 0.79
```



Train (90%)



Test (10%)



Only pre-training

- Micro_F1 : 12.7119
- AUPRC : 17.1305

With Fine-tuning

- Micro-F1 : 57.2581
- AUPRC : 60.4493

KLUE Benchmark

- Micro-F1 : 66.44
- AUPRC : 66.17

Hyperparameters

- Epochs : 3
- Batch Size : 2
- Learning Rate : 2e-5
- Evaluation Step : 60

- `anm:origin` 범위가 너무 넓었다.
- `anm:physical`과 `anm:disease`를 나눴으면 더 좋은 분포를 가졌을 것 같다.
- 모두가 만족하는 가이드라인은 없다.
- 부드럽고 명확한 커뮤니케이션은 정말 중요하다.
- Data imbalance 해소는 쉽지 않다.
- 데이터 제작은 상상 이상으로 더 많은 시간, 노력, 정신력이 소모된다.
(KLUE 제작하신 분들... 존경합니다...)

	좋았던 점	아쉬웠던 점
임동진	지금까지 No-Relation이 항상 가장 많았는데 왜 그럴 수 밖에 없는지 이해하게 되었습니다. 실제로 데이터 제작을 해보니 해당 Relation이 어떠한 이유로 데이터 형식을 이렇게 만들었고, 모델을 만들 때 어떤 방식으로 접근해야 할지 느낌이 온 것 같아 데이터를 읽는 것에 대한 중요성을 알게 되었습니다.	데이터를 많이 읽어본 적이 없어서 데이터 기준을 KLUE로만 잡았던 것이 아쉽습니다. 또한 anm:number_of_members를 최대한 늘려보려고 했으나 데이터의 한계로 많이 만들지 못했던 점이 아쉬웠습니다.
정재윤	이전대회 데이터셋에서 no_relation이 유독 많던 이유를 알 수 없었는데, 실제로 데이터 제작을 해보니 데이터에 대해 더 깊게 이해할 수 있었고, 데이터를 다루는 것만이 아닌 제작에 대한 경험을 쌓을 수 있어서 좋았습니다.	적지 않은 양을 직접 태깅을 하다보니 더 완벽한 데이터를 만들지 못함이 아쉬움으로 남았습니다.
조설아	데이터를 직접 제작하면서 데이터에 대한 가이드라인의 중요성을 깨닫게 되었습니다. 또한 entity type과 그에 따른 relation을 제한적으로 선정하는 과정을 통해, 선택과 집중이 어떤 방향으로 이뤄져야 하는지도 어렵듯이 감을 잡게 된 것 같아 매우 유의미한 경험이었습니다.	시간이 제한적이었던 점이 아쉬웠습니다. 특히 유연한 데이터 가이드라인이 오히려 어려움으로 다가와 이에 대한 확신이 부족했습니다. 피드백을 여러 번에 걸쳐 받았으면 좋았겠다는 아쉬움이 있었습니다.
허치영	데이터 제작 프로세스를 전반적으로 경험할 수 있는 기회를 얻어 명확한 구분 기준과 정의의 중요함에 대해 알 수 있었습니다. 커뮤니케이션시 조심해야하는 부분에 대해 깨달을 수 있었으며 소통 능력 중 부족한 부분을 다시 한번 알 수 있었습니다.	원시 데이터와 태스크의 분석에 훨씬 더 많은 노력과 시간을 들여 깊은 이해를 바탕으로 데이터를 제작했다면 더 좋은 결과물이 나오지 않았을까 하는 생각이 들어 아쉽습니다. 어느정도 균형 잡힌 데이터를 만들고 싶었으나 불균형한 분포를 나타내 아쉬웠습니다.
이보림	데이터 제작 프로세스를 경험하고 나니 데이터의 이해 및 소중함과 다양한 의견을 합일하는 방법에 대해 배웠다.	약간의 데이터 불균형을 보인 부분이 아쉬웠다.

경청해주셔서
감사합니다

QnA