

# 관계 추출 태스크 가이드라인

본 가이드라인은 관계 추출(Relation Extraction) 태스크를 위한 데이터 제작 과정을 설명하기 위해 작성되었습니다. Annotation에 참여하기 전에 반드시 숙지해주시기 바랍니다.

## [1. 관계 추출 태스크 알아보기](#)

## [2. 엔티티 상세 설명](#)

[ORG](#)

[PER](#)

[ANM](#)

[NOH](#)

[LOC](#)

[DAT](#)

[POH](#)

## [3. Annotation 가이드라인](#)

[3-1\) 관계의 방향성](#)

[3-2\) 관계 없음](#)

[3-3\) 외부 지식의 활용](#)

[3-4\) 관계의 시점](#)

## [4. Annotation 환경](#)

[4-1\) 관계 클래스 선택지](#)

## [5. FAQ](#)

---

## 1. 관계 추출 태스크 알아보기

관계 추출 태스크는 하나의 문장에서 나타나는 Entity 쌍 사이의 의미적 관계를 분류하는 태스크입니다. 반드시 문장은 하나만 주어지며, 이 문장에서 나타나는 Entity 쌍은 관계의 주체가 되는 Subject Entity와 대상이 되는 Object Entity로 이루어집니다.

Entity 쌍 사이에 나타날 수 있는 관계는 총 1개의 클래스 중 하나로 분류합니다. 관계 클래스는 구체적으로 1) 단체 중심 관계 1개, 2) 동물 중심 관계 8개, 3) “관계\_없음” 1개로 구성됩니다.

- 1) 단체 중심 관계는 Subject Entity가 기관, 회사, 정당, 국가 등의 단체일 때 발생할 수 있는 관계로서 “조직:하위조직”이 이에 해당합니다.
- 2) 동물 중심 관계는 Subject Entity가 동물일 때 발생할 수 있는 관계로서 “동물:서식지”, “동물:신체적 특징”, “동물:사냥감” 등이 이에 해당합니다.
- 3) “관계\_없음”은 주어진 문장에서 Entity 쌍이 아무 관계가 없음을 의미합니다. (3-2) [관계 없음](#) 에서 자세히 설명)

관계 추출 태스크에 대한 이해를 돕기 위해 다음의 예시를 살펴보겠습니다.

#### Example #1:

**Object** 오리너구리 의 전기수용 감각은 **Subject** 단공류 가운데서도 가장 뛰어나고 예민하다.

위의 문장에서 Entity 쌍을 이루는 Subject Entity는 “단공류”이며, Object Entity는 “오리너구리”입니다. 또한 위 문장에서 “오리너구리”와 “단공류” 사이의 의미적 관계는 “조직:하위 조직” 클래스라고 분류할 수 있습니다.

---

## 2. 엔티티 상세 설명

### 1. ORG

A는 개체수가 감소했으며, 인간에 의한 영향을 받고 있는 것이 명확해  
세부등급은 A1으로 지정됐다.

- 개체수 관련 관계를 지정하거나, 종속 관계를 지정하는 집단 및 개체
- 또한 동물 분류 체계에 해당하는 명칭, 동물의 품종 및 등급에 해당하는 명사
- 서식지나 경유지로 사용되지 않는 하나의 지명 혹은 단체

## 2. PER

봄베이고양이는 흑표범을 닮은 집고양이를 만들 목적으로 1958년 미국 켄터키 루이빌에서 **니키 호너(Nikki Horner)**가 버미즈와 아메리칸쇼트헤어를 교배시켜 만든 종이다.

- 사람의 이름, 직위 혹은 두 가지 모두 사용되는 경우
  - 예시 : 크루엘라 드 빌, 타이후르 왕, 왕자 등

## 3. ANM

**너구리**는 32종의 기생충의 숙주다.

- 독립적으로 서술된 동물을 ANM으로 지정
- 단, 너구리가 낳는 알 등의 생물이 아닌 부산물은 POH로 취급

## 4. NOH

봄베이고양이의 체중은 **2.5-5.4 킬로그램** 정도이며, 대개 수컷이 암컷보다 더 무겁다.

- 동물의 마릿수, 개수, 기온, 길이 혹은 무게의 단위

## 5. LOC

이 후 A는 아라비아 상인들의 **실크로드**를 통해 유럽과 아시아 전역으로 퍼져나갔다.

- 지역, 지명이라 하더라도 다른 Entity type으로 명확히 구분 가능한 경우를 제외하면 LOC로 지정

## 6. DAT

또한 레서판다는 수명이 평균 **8년** 정도 되는데, 새끼를 낳고, 기르는 것을 싫어해서 새끼를 낳지 않고 죽는 경우가 많다.

- 기간이나 특정 시점에 관련된 엔티티 타입

## 7. POH

개는 **사냥**, **목축**, **운송**, **경비**와 같은 일에 사용되고 있으며, 반려동물로 기르기도 한다.

- 위의 엔티티 타입들에 속하지 않는 엔티티 타입
- 이와 같은 예시에서는 종속관계가 없는 용도
- 신체 부위, 식물, 성분, 질병 및 역할에 관련된 명사

---

## 3. Annotation 가이드라인

### 3-1) 관계의 방향성

Subject에 대한 Object의 관계와 Object에 대한 Subject의 관계는 다를 수 있습니다. 즉, 경우에 따라 관계의 방향성이 존재합니다. 다음의 예시를 통해서 자세히 알아보겠습니다.

### Example #2:

**Subject**  
오리너구리

**Object**  
단공류

의 전기수용 감각은 가운데서도 가장 뛰어나고 예민하다.

Example #1과 다르게, 이번에는 “오리너구리”가 Subject, “단공류”가 Object Entity입니다. 이 경우에 관계는 “조직:하위 조직”가 아닙니다. “오리너구리”의 상위 종은 “단공류”가 아니기 때문입니다. 이 문장에서 “오리너구리”는 “단공류”를 포함하는 더 넓은 개념의 분류입니다. 따라서 이때의 관계는 “조직:하위 조직”이 적당하나, 현재 우리의 Relation에는 존재하지 않으므로 “관계\_없음”이 적합할 것입니다. 즉, 관계에는 방향성이 존재할 수 있고 동일한 Entity 조합이여도 Subject와 Object가 무엇이냐에 따라서 관계는 완전히 달라질 수 있습니다.

### 3-2) 관계 없음

문장에서 Entity 사이에 의미적 관계를 유추할 수 없거나, 정해진 관계 클래스에 존재하지 않을 수 있습니다. 이러한 경우, 주어진 샘플을 “관계\_없음” 클래스로 분류합니다.

### Example #3:

**Object**  
고양이

**Subject**  
사람

의 야간 시력은 보다 우수하나 낮 시간의 시력은 사람보다 열악하다.

위의 문장에서 “고양이”와 “사람”은 단순한 비교 대상일 뿐 어떠한 관계도 유추할 수 없습니다. 따라서 “관계\_없음”클래스로 분류합니다.

또한 Example #2처럼 관계가 실제로 존재하지만 Relation Map에 존재하지 않는 경우도 “관계\_없음” 클래스로 분류합니다.

### 3-3) 외부 지식의 활용

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.

#### Example #4:

Object 진도에서는 일반적으로 흑색이나 검붉은 색, 그리고 희고 검은 Subject 진돗개를 사냥개로서 좋다고 판단한다.

위 문장에서 “진돗개”와 “진도” 사이의 관계로 “동물:서식지” 등을 생각하기 쉬운데, 사실 이들 사이에는 관계가 없습니다(“관계\_없음”). 만약 “동물:서식지”라고 생각했다면 그것은 위의 문장 내에서 주어진 정보만으로 추론한 것이 아니라, “진도”라는 지역이 “진돗개”의 원서식지라는 사실, 외부 지식을 활용해서 유추한 것입니다. 위 문장에서는 “진도”라는 지역이 특정 형태의 “진돗개”를 사냥개로서 좋다고 판단했다는 정보만 나타날 뿐, 이를 통해서 “동물:서식지”의 관계를 유추하는 것은 불가능합니다. 따라서 위의 예시는 “관계\_없음”으로 분류합니다.

### 3-4) 관계의 시점

“동물:서식지”, “동물:개체수”, “동물:역할” 등의 관계들은 시간에 따라서 달라질 수 있습니다. 이런 관계의 경우 문장의 시점을 기준으로 관계가 존재하는지 판단합니다. 현재 시점이 아닌 문장의 시점을 기준으로 한다는 점에 주의합니다.

#### Example #5:

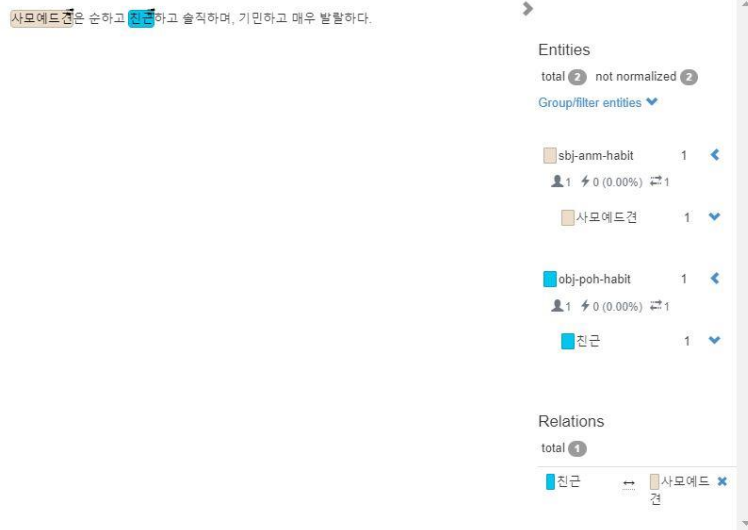
고양이의 조상은 Object **사막**에서 살았기 때문에 그 습성이 지금의 Subject **고양이**에게 전달된 것이다.

위 문장에서 “**고양이**”는 조상이 “**사막**”에 서식했기 때문에 문장의 시점을 기준으로 봤을 때 서식지가 아닙니다. 물론 “**고양이**”가 “**사막**”에 서식할 수는 있지만 이를 주어진 문장에서 유추하는 것은 불가능합니다. 따라서 위의 예시는 “관계\_없음”으로 분류합니다.

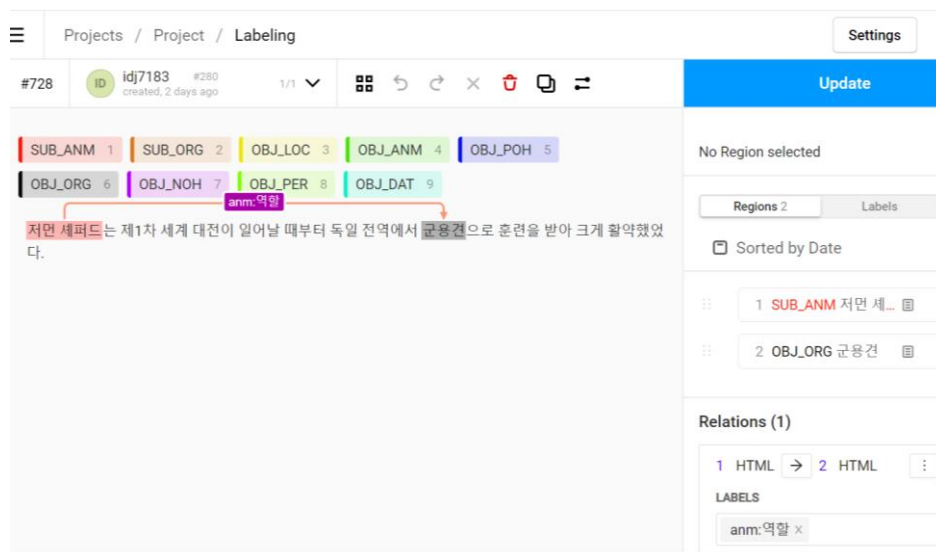
## 4. Annotation 환경

작업자 화면 예시:

### 1. Tagtog 환경



### 2. Label Studio 환경



작업자의 화면에는 위와 같이 1) 하나의 문장, 2) Subject/Object Entity, 3) 관계 클래스 선택지 (추천/일반)가 주어집니다.



#### 4-1) 관계 클래스 선택지

| sentence     | sentence_with_entity | subject_entity           | object_entity           | class                 |
|--------------|----------------------|--------------------------|-------------------------|-----------------------|
| 아프리카들고양이는 본래 | <obj:아프리카들고양이>는      | {'start': 13, 'end': 20, | {'start': 0, 'end': 10, | no_relation           |
|              |                      |                          |                         | anm:physical          |
|              |                      |                          |                         | anm:habit             |
|              |                      |                          |                         | anm:alternate_name    |
|              |                      |                          |                         | anm:habitat           |
|              |                      |                          |                         | anm:enemy             |
|              |                      |                          |                         | anm:role              |
|              |                      |                          |                         | anm:origin            |
|              |                      |                          |                         | anm:number_of_members |
|              |                      |                          |                         | org:sub_group         |
|              |                      |                          |                         | no_relation           |

가이드라인에 기반하여 주어진 문장에서 Subject와 Object Entity 사이의 관계를 주어진 관계 클래스 선택지 중에서 선택합니다. 추천 선택지는 주어진 샘플의 관계 클래스로 가능성이 높을만한 선택지로 구성되며, 추천 선택지에 적절한 것이 없을 경우 일반 선택지까지 확장하여 적절한 관계를 선택합니다. 만약 일반 선택지에도 적절한 것이 없다면 추천 선택지 최하단에 위치하는 “관계\_없음”을 선택합니다.

## 5. FAQ

Q. 기생충, 진드기 등의 생물을 ANM라고 생각해야 할까요?

- 아닙니다. 어디까지나 ANM은 “동물”에 해당하는 엔티티로 동물이 아닌 생물에 속하는 기생충, 진드기 등은 ANM로 생각할 수 없습니다.
- 기생충, 진드기 등은 POH 엔티티로 생각하시면 됩니다.

Q. POH와 ORG의 차이를 잘 모르겠습니다.

- 간단히 설명하자면 종속 관계에 존재하는가 존재하지 않는가에 의해 결정됩니다.
- 예를 들어, 아래와 같은 문장이 있다고 생각해봅시다.
  - 예시 문장 : 개는 사냥(POH), 목축(POH), 운송(POH), 경비(POH)와 같은 일에 사용되고 있으며, 반려동물(ORG)로 기르기도 한다.
  - 사냥, 목축, 운송, 경비는 “개”가 활용되는 분야이기는 하지만, 개가 사냥에 종속 관계인 것은 아닙니다. 따라서 POH Entity가 될 것입니다.
  - 하지만, 개는 “반려동물”이라는 범주에 속하는 동물입니다. 따라서, “반려동물”은 ORG Entity를 가지게 될 것입니다.

Q. 특정 장소를 LOC와 ORG 중 무엇으로 지정해야하는지 잘 모르겠습니다.

예시 문장 : 서울대공원에서 키우던 이들 6마리 풍산개의 2세들은 2001년 대한민국 각 지방 동물원에 분양되었으며, 이 후 일반인에게도 분양이 이루어졌다.

- 특정 장소에 동물이 서식한다는 말이 등장할 경우 LOC으로, 그외의 경우는 ORG으로 판단합니다.
- 장소로 해석될 여지가 있는 단어가 행위의 주체가 되었을 때, 그 문장에서는 장소보다는 단체로써 역할을 했다는 의미를 가집니다. 따라서 이 경우에는 ORG로, 나머지 경우에는 LOC로 보시면 될 것 같습니다.

Q. 어디까지 동물의 서식지, 즉 “동물:서식지”으로 보야 할까요?

예시 문장 : 보통 땅 위에서 생활하고, 땅 속이나 물 속에서 생활하는 종도 몇몇 있고, 나무 위에서 거의 대부분의 시간을 보내는 종도 있다.

- 거주에 동물의 의지가 들어가 있지 않거나, 영역을 의미하는 경우가 아니더라도 현재 그 장소에 살고 있다면 “동물:서식지” 관계로 취급합니다
- 즉, 자유의지로서의 서식(지명/자연 환경), 타의로서의 양식, 사육 등을 모두 포함하여 **현재 서식하고 있는지의 유무**만으로 판단합니다.
- 또한 “잡혔다”, “발견되었다” 등 직접적으로 서식하고 있다는 단어가 나오지 않는다면 “동물:서식지”으로 취급하지 않습니다.

Q. Entity에 용도가 들어가는 경우에는 사전지식을 활용한 것으로 보아야 하나요?

예시 문장 : 시베리안 허스키는 동시베리아에서 유래된 중형견으로 촘촘하게 덮인 모피를 지닌 썰매견이다.

- 이때, 시베리안 허스키는 “썰매를 모는”, “썰매용의”, “썰매용으로 사용된다” 등의 표현이 없어도, “썰매”라는 단어를 통해 썰매의 용도로 쓰인다는 것을 파악할 수 있습니다. 비슷한 예로는 “사냥개”, “구조견”등이 있습니다.

Q. 종/속/과 등을 사전지식으로 보아도 될까요?

- “속한다”/”중 하나이다”/”떨린” 등의 소속/포함 표현 없이 '종속과목강문계'만 언급된다면 “관계\_없음” 이라고 판단하였습니다.

Q. “동물:유래”의 범주가 어디까지인지 잘 모르겠습니다.

예시 문장 : 현대의 래브라도 종은 캐나다의 뉴펀들랜드 래브라도주(Newfoundland and Labrador)의 일부인 뉴펀들랜드섬에서 기원하였다.

- 역사/설화/속담/이야기/유래 등의 과거 기록에 등장한 인물, 사건, 지역 등을 모두 “동물:유래”로 판단하였습니다.
- 단, 문화에 관련된 이야기는 시간에 관계 없이 “동물:유래”로 취급하였습니다.

Q. 신체적/ 비신체적 특징의 구분 방법이 모호합니다.

“동물:신체적 특징”

- 원인을 해부학적으로 측정 가능한 지표 등에 대해서는 모두 “동물:신체적 특징”이라고 생각하였습니다.
- 또한 존재하는 기관에 의해 발생하는 모든 현상들도 “동물:신체적 특징”이라고 생각하였습니다. 예를 들어, 시각 같은 경우는 눈에 의해 발생하는 현상, 체온은

신체에서 에너지를 처리하며 발생하는 현상이기 때문에 "동물:신체적 특징"로 처리하였습니다.

#### "동물:비신체적 특징"

- 소위 동물의 습성이라 일컬어지는 특성들을 "동물:비신체적 특징"으로 보았습니다.
- 해부학적으로 명확한 판단이 불가능한 경우, 즉 일정 시간을 소요하여 판단할 수 있는 산란 장소, 수명 등의 요소를 "동물:비신체적 특징"으로 보았습니다.
- 특히 일생동안 일정하게 지니는 특성을 "동물:신체적 특징"로, 특정 시점/상황 등에서 변화가 있는 경우를 "동물:비신체적 특징"으로 보았습니다.
- 더불어 성격, 표정 등 사람의 주관에 따른 특성을 "동물:비신체적 특징"으로 판단하였습니다.
- 행동을 설명하는 신체 부위 또한 "동물:비신체적 특징"에 포함시켰습니다.
- 다른 동물 혹은 사람과의 관계도 해당 object의 특성, 즉 "동물:비신체적 특징"이라 보았습니다.
- '싫어한다'와 같은 부정 표현도 "선호"라는 특성으로 보아 "동물:비신체적 특징"에 포함하였습니다. 다만 부재의 표현은 포함시키지 않았습니다.

Q. 어떤 역할이 간접적으로 등장하였을 때는 어떤 관계로 봐야하나요?

예시 문장 : 영국에서는 달마시안을 코치도그라고 하여 마차를 탈 때나 승마할 때, 또는 여행의 안전을 위하여 데리고 다녔다.

- 직접적이든 간접적이든 관계 없이 해당 문장에서 목적성을 파악할 수만 있다면 "동물:역할"로 판단하고, 이외의 경우 "관계\_없음"으로 판단하기로 하였습니다.
- "역할"과 같은 직접 표현, "위하여"같은 간접 표현을 모두 포함하여 "동물:역할"을 설명하는 문구로 보았습니다.

Q. 사람이 주체가 되는 반려동물에 대한 유의사항, 관리법(ex. 씻기는 주기, 밥 주는 횟수 등)은 "동물:비신체적 특징"으로 봐야할까요? 아니면 "관계\_없음"으로 봐야할까요?

예시 문장 : 금붕어에게 먹이는 2개씩 매일 세 번씩 아침 점심 저녁으로 준다.

- 예외없이 "관계\_없음"으로 구분하고 있습니다.

Q. 존재하지 않는 것들은 특징으로 봐야하나요?

예시 문장 : 스팅크스처럼 털이 거의 없거나 멍크스처럼 꼬리가 없는 품종도 있다.

- 존재하지 않는 것들에 대해서는 특징으로 보지 않았습니다.
- 드물다, 거의 없다 등의 표현도 부재와 유사한 의미로 판단하였습니다.

Q. 직접적이 아닌 간접적으로 특징을 표현하였을 때 명시된 물체는 어떤 관계를 가지나요?

예시 문장 : 도롱뇽의 일생은 개구리와 같은 타 양서류와 비슷하다.

- 그 동물에 대해 직접적으로 설명하고 있지 않고, 만약 그 동물에 대해서 모르고 있다면 해당 문장에서는 관계를 알 수 없다고 생각하여 "관계\_없음"으로 판단하였습니다.

Q. 멸종 위기 등의 그 생명체가 원하지는 않았지만 특징이 된 것도 비신체적 특징으로 봐야할까요?

예시 문장 : 동네마다 흔하던 삽살개는 일제 강점기인 1940년 이후 일본이 개를 전쟁에 필요한 가축 공급원으로 삼아 많은 수를 죽임에 따라 그 수가 급격히 줄어 멸종의 위기에까지 이르렀다.

- 멸종위기는 그 동물이기 때문에 발생한 문제라고 생각하였습니다. 따라서, 생명체가 원하지 않았더라도 그 생명체이기 때문에 발생한 것들에 대해서는 "동물:비신체적 특징"으로 판단하기로 했습니다.
- Outlier도 같은 맥락에서 특징으로 판단하기로 했습니다.