

태양계의 형성과 진화

관계 추출 태스크 가이드라인

목차

[목차](#)

[1. 작업 목적](#)

[2. 텍스트 데이터 전처리](#)

[3. Entity](#)

[3.1\) PER \(PERSON\)](#)

[3.2\) CLO \(CELESTIAL OBJECT\)](#)

[3.3\) CPT \(CONCEPT\)](#)

[3.4\) DAT \(DATE\)](#)

[3.5\) ELM \(ELEMENT\)](#)

[3.6\) MET \(METRIC\)](#)

[4. Relation](#)

[4.1 관계의 종류](#)

[4.1.1 no_relation](#)

[4.1.2 clo:revolves](#)

[4.1.3 clo:exists_in](#)

[4.1.4 clo:contains](#)

[4.1.5 clo:turn_into](#)

[4.1.6 clo:alias_of](#)

[4.1.7 clo:composed_of](#)

[4.1.8 met:feature_of](#)

[4.1.9 per:propose](#)

[4.1.10 dat:date_of_discovery](#)

[4.1.11 per:origin_of](#)

[4.2 관계의 방향성](#)

[4.3 외부지식의 활용](#)

[4.4 관계의 시점](#)

[5. Annotation 환경](#)

[5-1\) 관계 클래스 선택지](#)

[5-2\) Data Error](#)

[5-3\) Hate / Bias / Privacy](#)

[6. FURTHER QUESTION](#)

1. 작업 목적

관계 추출 태스크는 한 개 또는 두 개의 문장에서 나타나는 Entity 쌍 사이의 의미적 관계를 분류하는 태스크다. 문장에서 나타나는 Entity 쌍은 관계의 주체가 되는 Subject Entity와 대상이 되는 Object Entity로 이루어진다.

Entity 쌍 사이에 나타날 수 있는 관계는 총 11개의 클래스 중 하나로 분류한다. 관계 클래스는 구체적으로 1) 천체 중심 관계 6개, 2) 인물 중심 관계 2개, 3) 수치 중심 관계 1개, 4) 날짜 중심 관계 1개, 5) “관계_없음” 1개로 구성된다.

1. 천체 중심 관계는 Subject Entity가 천체일 때 발생할 수 있는 관계로서 “천체:공전하다”, “천체:위치하다”, “천체:포함하다”, “천체:변화하다”, “천체:이루어지다”, “천체:별칭이다”가 이에 해당한다.
2. 인물 중심 관계는 Subject Entity가 인물일 때 발생할 수 있는 관계로서 “인물:제안하다”, “인물:기원이 되다”가 이에 해당한다.
3. 수치 중심 관계는 Subject Entity가 수치일 때 발생할 수 있는 관계로서 “수치:특성이다”가 이에 해당한다.
4. 날짜 중심 관계는 Subject Entity가 날짜일 때 발생할 수 있는 관계로서 “날짜:발견한 날짜” 등이 이에 해당한다.
5. “관계_없음”은 주어진 문장에서 Entity 쌍이 아무 관계가 없음을 의미한다. (2-2) [관계 없음](#) 에서 자세히 설명)

2. 텍스트 데이터 전처리

전체 텍스트 세트 목록은 다음과 같다.

- 고리, 과학, 궤도, 금성, 달, 명왕성, 목성, 물리학, 복사, 블랙홀, 성간, 성운, 소행성, 수성, 수소, 안드로메다, 온도, 왜성, 위성, 은하, 이심률, 중력, 지구, 지질학, 진화, 천문학, 천왕성, 초신성, 충돌 태양, 태양계, 토성, 항성, 해왕성, 핵융합, 행성, 헬륨, 화성

본 과제에서 추출할 **relation**은 천문학과 관련된 텍스트로 제한한다.

주어진 문장 내에서 주체와 대상 **Entity**를 찾을 수 없거나, 주제와 맞지 않는 텍스트 세트는 제외한다. 제외할 텍스트 세트 목록은 다음과 같다.

- 이심률, 복사, 안드로메다, 성간, 궤도, 고리, 진화, 지질학, 충돌

명확한 태깅 작업을 위해서 각 텍스트 세트 내 문장들을 아래와 같은 기준으로 처리한다.

1. 문장은 ‘.’을 기준으로 나눈다.
2. 문장을 분리하거나 합쳤을 때, 기존 문장보다 주체와 대상의 관계가 더 명확해지는 경우에는 문장을 분리하거나 합친다.
3. **Entity**를 태깅하기 분명하지 않거나 주관적 판단이 필요한 문장은 제거한다.
4. 문장에서 대명사가 **Entity**로 사용되는 경우, 대명사가 의미하는 구체적인 명사로 변경하고 변경이 불가능한 경우에는 해당 문장을 제거한다.
5. 문장 구조에서 주체나 대상이 생략된 경우 적절한 명사를 추가하여 **Entity**가 명확하게 정의되도록 한다.

3.Entity

전반적인 Entity 태깅 기준은 다음과 같다.

1. ‘디오(Dione)’와 같이 명사 뒤에 동일한 의미의 외국어가 함께 등장하는 경우 모두 Entity로 태깅한다.
2. 고유명사와 일반명사가 모두 등장할 경우 고유명사를 우선적으로 태깅한다.
3. 주체나 대상에 해당하는 Entity가 2개 이상 등장할 경우 가장 먼저 등장하는 Entity 하나만 태깅한다.
4. PER과 CLO로 분류가 되는 단어는 CPT로 분류 될 수 없다.

관계추출 테스트에 등장하는 Entity는 다음 Entity 중 하나여야 한다.

3.1) PER (PERSON)

PERSON은 사람을 의미한다. 본 과제에서는 사람 이름과 신 이름으로 제한한다.

- Example 1

1915년, 알베르트 아인슈타인이 일반상대론을 고안하여 중력이 빛의 운동에 영향을 미침을 보였다.

“알베르트 아인슈타인”은 사람 이름이다. 따라서 PER Entity라 할 수 있다.

- Example 2

해왕(海王)은 ‘바다의 왕’이라는 한자어로, 포세이돈(그리스 신화명) 또는 넵투누스(로마 신화명)를 번역한 것이다.

“포세이돈”과 “넵투누스”는 신의 이름이다. 따라서 PER Entity라 할 수 있다.

“해왕(海王)”과 “바다의 왕”은 사람이나 신의 이름이 아니다. 따라서 PER Entity라 할 수 없다.

3.2) CLO (CELESTIAL OBJECT)

CELESTIAL OBJECT는 천체를 의미한다. <https://ko.wikipedia.org/wiki/천체>의 천체, 가스 행성,떠돌이 별, 불박이별

- Example 1

항성은 수소 및 헬륨, 기타 중원소로 이루어진 성간 구름이 붕괴하면서 탄생한다.

“항성”과 “성간 구름”은 천체이다. 따라서 CLO Entity라 할 수 있다.

- Example 2

식쌍성, **흑점**이 매우 많은 별(자전하면서 밝기가 변화한다.) 등이 그 예이다.

“식쌍성”은 관찰자의 시점에서 볼 때 두 **항성**의 궤도면이 아주 가까워서 서로 식 현상이 발생하는 쌍성이고, 쌍성은 항성계에 속한다. 따라서 **CLO Entity**라 할 수 있다.

“흑점”은 태양의 광구에 존재하는 영역이다. 영역은 천체가 아니다. 따라서 **CLO Entity**라 할 수 없다.

- Example 3

높은 고도에서 **해왕성 대기**는 80%가 수소, 19%가 헬륨이다.

“해왕성”은 태양계에 속하는 행성이다. 따라서 **CLO Entity**라 할 수 있다.

“해왕성 대기”는 해왕성의 주위를 대체로 일정하게 둘러싸고 있는 기체층이다. 기체층은 천체가 아니다. 따라서 **CLO Entity**라 할 수 없다.

- Example 4

예를 들면 외계행성 **<HD 209458 b>**의 질량은 0.69 "MJ"이며, 안드로메다자리 **<카파 b>**의 질량은 12.8 "MJ"이다.

“HD 209458 b”, “카파 b”은 행성의 이름이다. 따라서 **CLO Entity**라 할 수 있다.

3.3) CPT (CONCEPT)

CONCEPT은 법칙, 이론, 개념을 의미한다. 이때 책 이름과 같이 특정 개념을 나타내는 것이 아닌 경우는 제외한다. 또한 위키피디아에 문서로 등장하지 않는 경우 제외한다.

- Example 1

우주 마이크로파 배경, **허블의 법칙**, 우주의 원소 함량 등의 관측이 지지하는 **대폭발 이론**의 등장으로, **물리적 우주론**은 20세기 들어와 큰 성공을 거두었다.

“허블의 법칙”, “대폭발 이론”, “물리적 우주론”은 법칙이나 이론이다. 따라서 **CPT Entity**라 할 수 있다.

- Example 2

그는 고리 사이의 검은 선으로 보이는 거대한 간격을 찾아냈으며, 이 간격이 바로 **'카시니간극'**(카시니 틈, **Cassini division**)'이다.

“카시니간극”은 고리 사이의 검은 선으로 보이는 거대한 간격을 의미하는 단어다. 따라서 **CPT Entity**라 할 수 있다.

- Example 3

블랙홀로 특징지어지는 **일반상대론의 최초의 근대적 해**는 1916년 카를 슈바르츠실트가 발견했다.

“일반상대론의 최초의 근대적 해”는 위키피디아에 문서로 등장하지 않는다. 따라서 **CPT Entity**라 볼 수 없다.

3.4) DAT (DATE)

DATE는 시기를 의미한다. “1920년”, “19세기” 등 특정 시점이나 “45년 전”과 같이 특정 시기를 의미하는 단어로 제한한다.

- Example 1

1923년에는 1000번째, 1990년에는 5000번째 소행성이 발견되었으며, 2013년 1월 30일 기준 35만3926개의 소행성에 공식적으로 숫자가 부여되었다.

“1923년”는 특정 시기를 의미한다. 따라서 DAT Entity라 할 수 있다.

- Example 2

지금으로부터 40억 년 전(태양계가 생겨나고 5~6억 년 뒤) 후기 대폭격이 있었는데 수성 표면에 있는 많은 충돌구(운석 구덩이, 크레이터)는 이 때 생겼다.

“40억 년 전”은 40억 년 전의 특정 시기를 의미한다. 따라서 DAT Entity라 할 수 있다.

“5~6억 년 뒤”는 그 자체로 특정 시기가 아니다. 따라서 DAT Entity라 할 수 없다.

- Example 3

또한 이들 계는 10 ~ 50년 동안 불과 수 개월 동안만 엑스선을 활발하게 방출한다.

“10 ~ 50년”은 기간을 의미한다. 기간은 특정시기로 볼 수 없다. 따라서 DAT Entity라 할 수 없다.

- Example 4

새를 제외한 지구상의 모든 공룡이 멸종한 백악기 제3기의 K-T 대멸종은 널리 알려져 있다.

“백악기 제3기” 특정 시기를 의미한다. 따라서 DAT Entity라 할 수 있다.

- Example 5

DAT:1980 - 1990년대, 나트륨, 칼륨과 칼슘이 수성 대기 속에서 발견됐다.

“1980 - 1990년대”는 특정 시기를 의미한다. 따라서 DAT Entity라 할 수 있다.

3.5) ELM (ELEMENT)

ELEMENT는 원소, 물, 암석 등의 물질을 의미한다. 대기와 같이 영역은 제외한다.

- Example 1

이것의 97%는 수소이고 3%는 헬륨이다.

“수소”와 “헬륨”은 원소다. 따라서 ELM Entity라 할 수 있다.

- Example 2

별의 축퇴압이 물질을 축퇴물질로 응축시키면 붕괴는 멈춘다.

“축퇴물질”은 페르미 입자가 페르미-디랙 통계를 따르기 위해 저온에서 나타내는 행동이다. 행동은 물리적인 요소가 아니다. 따라서 ELM Entity라 할 수 없다.

- Example 3

타이탄은 표면 중력이 작음에도 불구하고 온도가 낮아(약 -180°C) 짙은 대기를 가지고 있었는데, 1944년 천문학자 제러드 카이퍼는 타이탄의 대기에 메탄이 포함되어 있다는 것을 발견했다.

“대기”는 기체층을 의미한다. 기체층은 물질이 아니다. 따라서 대기는 ELE Entity라 할 수 없다.

3.6) MET (METRIC)

METRIC은 온도, 무게, 밀도, 자전주기 등 한 물체의 수치적인 특성을 의미한다. 숫자와 단위가 동시에 들어간 경우로 제한한다. 상대적인 수치는 제외한다.

- Example 1

수성은 태양계 행성들 중 가장 반지름이 작은 행성으로, 구체적인 크기는 2439.7km이다.

“2439.7km”는 숫자와 단위가 모두 등장하고 수성의 크기를 절대적 수치로 나타낸다. 따라서 MET Entity라 할 수 있다.

- Example 2

화성의 지름은 지구의 1/2정도이다.

“1/2”은 단위가 없고 상대적인 수치이다. 따라서 MET Entity라 할 수 없다.

- Example 3

지구형 행성을 이루는 물질은 우주에서 매우 희귀한 존재이기 때문에(성운 질량 0.6퍼센트에 불과하다.), 지구형 행성은 크게 자라날 수가 없었다.

“0.6퍼센트”는 단위는 있지만 상대적인 수치이다. 따라서 MET Entity라 할 수 없다.

- Example 4

궤도 이심률이 0.056이기 때문에 토성과 태양 사이 거리는 태양에 가장 가까울 때와 멀어졌을 때 차이가 약 155,000,000 킬로미터이다.

“0.056”은 단위가 없다. 또한 궤도 이심률이 두 천체에 관한 개념이기에 MET Entity라 할 수 없다.

“약 155,000,000 킬로미터” 두 천체의 거리에 대한 내용이기 때문에 MET Entity라 할 수 없다.

- Example 5

공전궤도 평균 속도는 9.69km/s로, 지구 시간으로 태양을 1회 공전하는 데에는 10,759일(29.5년)이 걸린다.

“9.69km/s”와 “10759일”은 두 천체의 공전에 관한 내용이기 때문에 MET Entity라 할 수 없다.

- Example 6

토성 궤도 경사각은 지구 공전면에 대해 2.48도 기울어져 있다

“2.48도”는 토성과 지구 두 천체에 관한 개념이기 때문에 MET Entity라 할 수 없다.

- Example 7

여기서 MET:1등급 별은 MET:2등급 별보다 2.5배 밝고, MET:6등급 별보다 100배 밝다는 것을 알 수 있다.

“1등급”, “2등급”, “6등급”은 숫자와 단위가 모두 들어간 수치이다. 따라서 MET Entity라 할 수 있다.

“2.5배”, “100배”는 상대적인 수치이다. 따라서 MET Entity라 할 수 없다.

- Example 8

태양의 겉보기 등급은 -26.7이지만 절대 등급은 +4.83에 불과하다.

“-2.67”, “+4.83”은 단위가 없다. 따라서 MET Entity라 할 수 없다.

4. Relation

4.1 관계의 종류

4.1.1 no_relation

Entity간의 관계를 정의할 수 없을 때 태깅한다.

4.1.2 clo:revolves

- Entity : <SUBJ-CLO> - <OBJ-CLO>
- <SUBJ-CLO>가 <OBJ-CLO>를 공전할 때 또는 <SUBJ-CLO>가 <OBJ-CLO>의 위성 혹은 행성일 때 태깅.
 - <OBJ-CLO: 토성> 주위를 불규칙하게 도는 <SUBJ-CLO: 히페리온>이나 ...
 - <SUBJ-CLO: 달>은 <OBJ-CLO: 지구>의 유일한 자연위성이고,
 - <SUBJ-CLO: 행성>이 <OBJ-CLO: 태양> 주변을 공전한다.
 - <SUBJ-CLO: 지구>가 <OBJ-CLO: 태양>의 행성이다.
 - <SUBJ-CLO: 위성>은 <OBJ-CLO: 행성> 따위의 둘레를 도는 천체를 말한다.
 - <OBJ-CLO: 마케마케>는 1개의 <SUBJ-CLO: 위성>을 가지고 있다.
- 안 되는 예시
 - <OBJ-CLO: 행성> 주위를 도는 사람이 만든 위성은 <SUBJ-CLO: 인공위성>이라고 한다.
:인공위성은 CLO(천체)가 아니라서 제외

4.1.3 clo:exists_in

- Entity : <SUBJ-CLO> - <OBJ-CLO>
- <SUBJ-CLO>가 <OBJ-CLO>에 위치하는 경우 태깅 (물리적 위치 기준)
 - <SUBJ-CLO: 명왕성>은 <OBJ-CLO: 카이퍼 대>에 있는 왜행성이다.
 - <SUBJ-CLO: 소행성>은 화성과 목성 궤도 사이의 <OBJ-CLO: 소행성대>에 존재한다.
 - <SUBJ-CLO: 토성>은 태양으로부터 여섯 번째에 있는 <OBJ-CLO: 태양계>의 행성이다.

4.1.4 clo:contains

- Entity : <SUBJ-CLO> - <OBJ-CLO>
- <SUBJ-CLO>의 개념이 <OBJ-CLO>의 개념을 포함하는 경우 태깅 (OBJ는 SUBJ이다 -> SUBJ는 OBJ이다는 아님)
 - <SUBJ-CLO: 성운>의 예는 <OBJ-CLO: 장미 성운> 또는 펠리칸 성운이다.
 - 처음으로 발견된 <SUBJ-CLO: 소행성>은 <OBJ-CLO: 세레스>이며,
 - 내행성계로 일컫는 <SUBJ-CLO: 지구형 행성>들로는 <OBJ-CLO: 수성>, 금성, 지구, 화성이 있다.
 - <OBJ-CLO: 천왕성>, 해왕성은 <SUBJ-CLO: 거대 얼음 행성>이다.

4.1.5 clo:turn_into

- Entity : <SUBJ-CLO> - <OBJ-CLO>
- <SUBJ-CLO>가 <OBJ-CLO>로 변화했을 경우 태깅 (SUBJ가 과거, OBJ가 미래)
 - <OBJ-CLO: 항성>은 수소 및 헬륨, 기타 중원소로 이루어진 <SUBJ-CLO: 성간 구름>이 붕괴하면서 탄생한다.
 - 가스로 된 코마나 꼬리가 없다는 점에서 혜성과 구분되지만, 일부 <OBJ-CLO: 소행성>은 과거에 <SUBJ-CLO: 혜성>이었다.
 - <OBJ-CLO: 블랙홀>은 <SUBJ-CLO: 항성>이 진화의 최종단계에서 폭발 후 수축되어 생성된 것으로 추측되는,

4.1.6 clo:alias_of

- Entity : <SUBJ-CLO> - <OBJ-CLO>
- <SUBJ-CLO>가 <OBJ-CLO>의 별칭일 경우 태깅 (OBJ는 SUBJ이다, SUBJ는 OBJ이다 둘 다 성립해야 함)
 - 또한 <SUBJ-CLO:토성>은 <OBJ-CLO: "Jewel of the Solar System">(태양계의 보석)>이라는 별명도 가지고 있다.
 - <SUBJ-CLO:행성(行星,)> 또는 <OBJ-CLO:혹성(惑星)>, 떠돌이별은 우주에서 항성의 둘레를 도는 천체의 한 부류이다.
 - 그들은 <SUBJ-CLO:수성>을 <OBJ-CLO:나부>라고 불렀다.

4.1.7 clo:composed_of

- Entity : <SUBJ-CLO> - <OBJ-ELM>
- <SUBJ-CLO>가 <OBJ-ELM>로 이루어졌을 때(구성되었을 때) 태깅
 - <SUBJ-CLO:태양>과 가스 행성(목성, 토성, 천왕성, 해왕성) 들도 <OBJ-ELM:수소>와 헬륨이 전체의 대부분을 차지하고 있다.
 - <SUBJ-CLO:금성>의 표면은 평탄한 <OBJ-ELM:현무암>질 평원으로 되어 있다.
 - <SUBJ-CLO:금성>은 <OBJ-ELM:이산화탄소>가 풍부한 대기에 의해서 온실효과가 발생,
 - <SUBJ-CLO:금성>의 핵은 지구와 마찬가지로 <OBJ-ELM:철>과 니켈이 주성분이다.

4.1.8 met:feature_of

- Entity : <SUBJ-MET> - <OBJ-CLO>
- <SUBJ-MET>가 <OBJ-CLO치>의 수치적 특성일 때(크기, 온도, 속도, 밀도 등) 태깅
 - <OBJ-CLO:수성> 평균 온도는 <SUBJ-MET:452.5K>이다.
 - <OBJ-CLO:수성>의 자전 주기는 <SUBJ-MET:58일>이다.
 - <SUBJ-MET:6등급> <OBJ-CLO:별>보다 100배 밝다는 것을 알 수 있다.
- 안 되는 예시
 - <OBJ-CLO:베토벤 분지>는 분출물 덮개와 비슷한 크기이며, 폭은 <SUBJ-MET:625km>이다.
: 베토벤 분지는 천체(CLO)가 아니라서 제외
 - <OBJ-CLO:달> 핵의 반지름은 <SUBJ-MET: 300~425km> 사이, 나머지 부분은 맨틀을 이루고 있다.
: 달이 아니라 <달 핵>인데, <달 핵>은 천체체(CLO)가 아니라서 제외
 - <OBJ-CLO:달>은 지구에서 약 <SUBJ-MET:38만 5천km> 떨어진 곳에 있으며,
: <달과 지구>에 대한 수치이지 달에 대한 수치가 아니라서 제외
 - <OBJ-CLO:금성>은 <SUBJ-MET:태양> 주위를 224일 주기로 돌고 있는데 반해, 자전 주기는 243일로 공전주기보다 길다.
: 자전 주기인 243일은 금성의 수치가 맞지만, 공전 주기 224일 금성과 태양 사이의 수치이므로 제외

4.1.9 per:propose

- Entity : <SUB-PER> - <OBJ-CON>
- 어떤 사람(PER)이 개념, 이론, 법칙, 현상 등을 제안/제시했을 때 태깅
 - 1915년, <SUBJ-PER:알베르트 아인슈타인>이 <OBJ-CON:일반상대론>을 고안하여 ...
 - 기원전 2세기에는 <SUBJ-PER:히파르쿠스>가 <OBJ-CON:세차>를 발견하였고 달의 크기와 거리를 계산하였으며,
 - 이탈리아의 천문학자 <SUBJ-PER:카시니>는 고리 사이의 검은 선으로 보이는 거대한 간격을 찾아냈으며, 이 간격이 바로 <OBJ-CON:'카시니간극'>이다.
- 안 되는 예시
 - 이 용어의 영단어 <OBJ-CON: 드워프 스타(dwarf star)>는 1906년 덴마크의 천문학자 <SUB-PER: 아이나르 헤르츠스프룽>이 하버드 스킴(Havard scheme)에서 K와 M으로 분류되는 가장 붉은 항성들이 2개의 구별되는 그룹으로 나눌 수 있다는 것을 발견한 1906년에 처음 만들어졌다.
: '드워프 스타'라는 단어를 '아이나르 헤르츠스프룽'이 만든 단어인지 명확하지 않기 때문에 제외
 - 1967년에 존 아치볼드 휠러가 강의에서 '블랙홀'이라는 말을 사용하여 일부에서는 휠러가 이 말을 고안했다고 보기도 한다.
: '보기도 한다.'는 확언이 아니기 때문에 제외
 - <OBJ-CON: '초신성'(supernova)>이라는 단어는 1931년에 <SUB-PER: 발터 바데>와 프리츠 츠비키가 만들어낸 조어이다.
: 초신성은 CONCEPT이 아닌 CLO이므로 제외

4.1.10 dat:date_of_discovery

- Entity : <SUBJ-DAT> - <OBJ-CLO>
- <SUBJ-DAT>가 <OBJ-CLO>의 발견 날짜일 때(년도, 세기, 연월일, ~년 전 등) 태깅
 - <SUBJ-DAT: 964년>에는 페르시아 천문학자 알 수피가 <OBJ-CLO: 안드로메다>를 발견하였다.
 - <SUBJ-DAT: 2016년 2월 11일>, LIGO 합동연구진은 두 개의 블랙홀이 서로 융합하면서 발생한 중력파를 감지함으로써 역사상 최초의 <OBJ-CLO: 중력파> 관측에 성공했다고 발표했다.
 - <OBJ-CLO: '초신성'(supernova)>이라는 단어는 <SUBJ-DAT: 1931년>에 발터 바데와 프리츠 츠비키가 만들어낸 조어이다.

- 안 되는 예시
 - <SUBJ-DAT: 1838년>에는 베셀이 <OBJ-CLO: 백조자리 61별>의 연주시차를 발견하였다.
: 연주시차를 발견한 것이지 백조자리 61별을 발견한 것이 아니기 때문에 제외
 - 이 용어의 영단어 <OBJ-CLO: 드워프 스타(dwarf star)>는 1906년 덴마크의 천문학자 아이나르 헤르츠스프룽이 하버드 스킴(Havard scheme)에서 K와 M으로 분류되는 가장 붉은 항성들이 2개의 구별되는 그룹으로 나뉠 수 있다는 것을 발견한 <SUBJ-DAT: 1906년>에 처음 만들어졌다.
: 무엇이 언제 누구에 의해 만들어지고 발견되었는지 확실히 알 수 없어서 제외

4.1.11 per:origin_of

- Entity : <SUBJ-PER> - <OBJ-CLO>
- <SUBJ-PER>가 <OBJ-CLO>이름의 기원 또는 유래일 때(PER은 사람 또는 신 이름)
 - <OBJ-CLO: Saturn>은 로마의 신 '<SUBJ-PER:사투르누스(Saturnus)>'에 기원한 것이다.
- 안 되는 예시
 - <OBJ-CLO: 성운>의 서양 명칭인 네불라는 구름을 의미하는 라틴어(<SUBJ-PER: nebula>)에서 유래되었다.
: nebula는 사람이 아니라서 제외
 - '신성'('"'Nova"'')이란 '새로운'이라는 의미의 라틴어 낱말에서 유래된 것으로, 천구상에 매우 밝은 별이 새로 나타난 것처럼 보이는 것을 칭한 것이며, 접두사 '초-'(super-)는 초신성이 광도가 훨씬 떨어지는 보통의 신성과는 구분되는 존재라는 것을 의미한다.
: '새로운'은 사람이 아니라서 제외

4.2 관계의 방향성

- clo:revolves, clo:exists_in, clo:contains, clo:turn_into, clo:composed of는 SUBJ와 OBJ가 바뀌면 성립하지 않기 때문에 주의해야 한다.
- clo:alias_of는 SUBJ와 OBJ가 바뀌어도 되기 때문에 먼저 등장하는 것을 SUBJ, 나중에 등장하는 것을 OBJ로 한다.
- met:feature_of, per:propose, dat:date_of_discovery, per:origin_of는 MET, PER, DAT를 SUBJ로, CLO를 OBJ로 한다.

4.3 외부지식의 활용

- 헬륨이라는 말은 그리스어로 태양을 뜻하는 말 헬리오스(Helios)에서 유래하였다.
-> 헬리오스는 태양신이라는 의미도 있지만 여기에서는 알 수 없기 때문에 PER로 태깅할 수 없다.

4.4 관계의 시점

- cls:turn_into의 경우 과거의 천체가 SUBJ, 현재의 천체가 OBJ로 태깅되기 때문에 주의해야 한다.
<OBJ-CLO: 항성>은 수소 및 헬륨, 기타 중원소로 이루어진 <SUBJ-CLO: 성간 구름>이 붕괴하면서 탄생한다.

5. Annotation 환경

작업자 화면 예시:

1915년, SUBJECT 알베르트 아인슈타인이 OBJECT 일반상대론을 고안하여
중력이 빛의 운동에 영향을 미침을 보였다.

관계 선택지

- | | | |
|-------------------------------------|--|---------------------------------|
| <input type="radio"/> no_relation | <input type="radio"/> clo:composed_of | <input type="radio"/> 데이터 Error |
| <input type="radio"/> clo:revolves | <input type="radio"/> met:feature_of | |
| <input type="radio"/> clo:exists_in | <input checked="" type="radio"/> per:propose | |
| <input type="radio"/> clo:conraints | <input type="radio"/> dat:date_of_discovery | |
| <input type="radio"/> clo:turn_into | <input type="radio"/> per:origin_of | |
| <input type="radio"/> clo:alias_of | | |

작업자의 화면에는 위와 같이 1) 한 개 혹은 두 개의 문장, 2) Subject/Object Entity, 3) 관계 클래스 선택지, 4) Data Error 체크박스가 주어집니다.

5-1) 관계 클래스 선택지

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.

5-2) Data Error

만약 데이터에 다음과 같은 오류가 있다고 판단되면 “Data Error” 체크박스를 눌러주세요.

- Entity 범위 오류
- 완전하지 않은 문장

Example #6:

1915년, SUBJECT 알베르트 아인슈타인이 일반OBJECT 상대론을 고안하여
중력이 빛의 운동에 영향을 미침을 보였다.

Entity 범위 오류: 위의 예시에서 Subject Entity는 의미적으로 보았을 때 “알베르트 아인슈타인”이지만 범위의 오류가 있어서 “알베르트 아인”으로 표시되고 있습니다. 또한 Object Entity 역시 의미적으로는 “일반상대론”이라는 것을 알 수 있지만 “상대론”까지 범위가 잘못 설정되어 있습니다. 이런 경우 의미적으로 나타나는 Entity를 기준으로 관계 클래스를 선택하고 “Data Error” 체크박스도 함께 눌러주시면 됩니다. 위 예시에서는 의미적으로 보았을 때, Subject Entity는 “알베르트 아인슈타인”, Object Entity는 “일반상대론”이며 이들의 관계로 “per:propose” 클래스를 선택하고 “Data Error” 체크박스를 누르면 됩니다.

Example #7:

SUBJECT 금성은 극도로 두꺼운 대기를 가지고 OBJECT 이산화탄소가 풍부한 대기에

완전하지 않은 문장: 위의 예시와 같이 Entity는 정상적으로 표시되고 있지만, 문장의 끝 부분이 잘려나간 것처럼 보입니다. 이와 같이 문장의 일부가 잘려나가거나 정상적인 문장으로 보이지 않는 경우 “Data Error” 체크박스를 누르면 됩니다. 이때도 마찬가지로 주어진 문장 내에서 최대한 관계를 유추해서 관계 클래스를 함께 선택해야 합니다.

Example #8:

불과 몇 개월 뒤, SUBJECT 카를 슈바르츠실트가 OBJECT 전지량과 구질량의 중력장을 기술하는 아인슈타인 방정식의 해를 구하였다(슈바르츠실트 계량)

완전하지 않은 문장: 위의 예시와 같이 Entity는 정상적으로 표시되고 있지만, 앞에 등장하는 “불과 몇 개월 뒤”는 하나의 문장에서 해결 할 수 없습니다. 이와 같이 완전한 문장으로 보이지 않는 경우 “Data Error” 체크박스를 누르면 됩니다. 이때도 마찬가지로 주어진 문장 내에서 최대한 관계를 유추해서 관계 클래스를 함께 선택해야 합니다.

5-3) Hate / Bias / Privacy

주제가 태양계이기 때문에 데이터에 혐오 표현, 편향적 표현, 개인정보가 포함되어 있지 않아서 “Hate / Bias / Privacy”에 대한 선택지는 따로 설정하지 않았다.