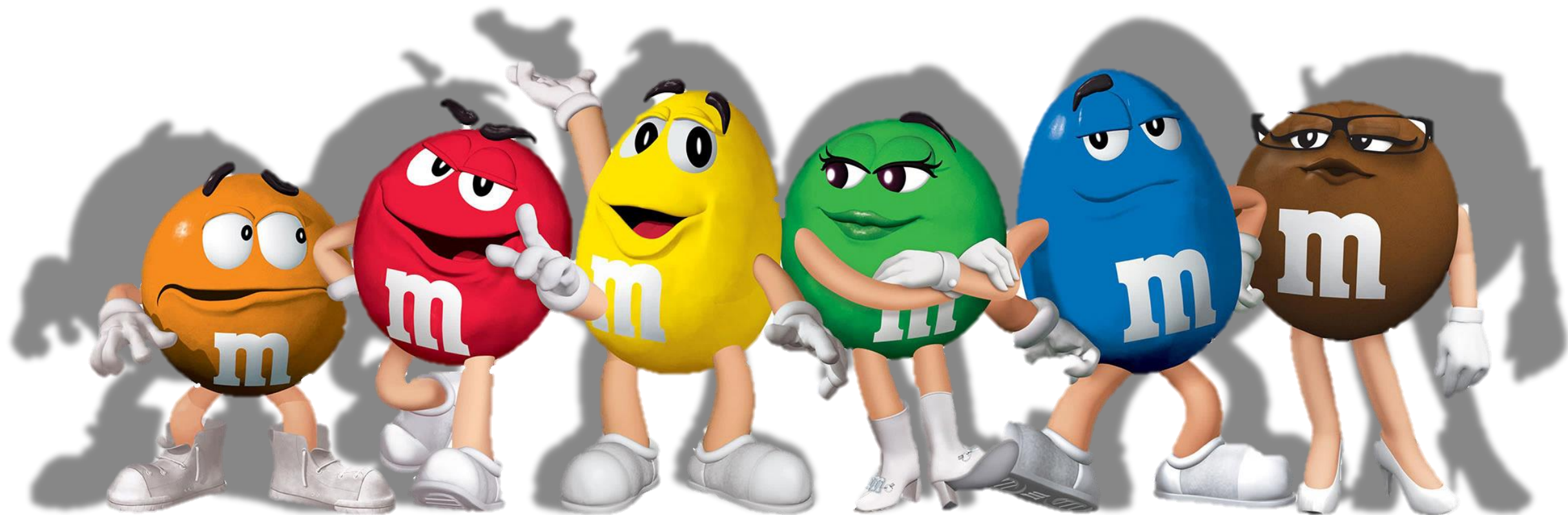
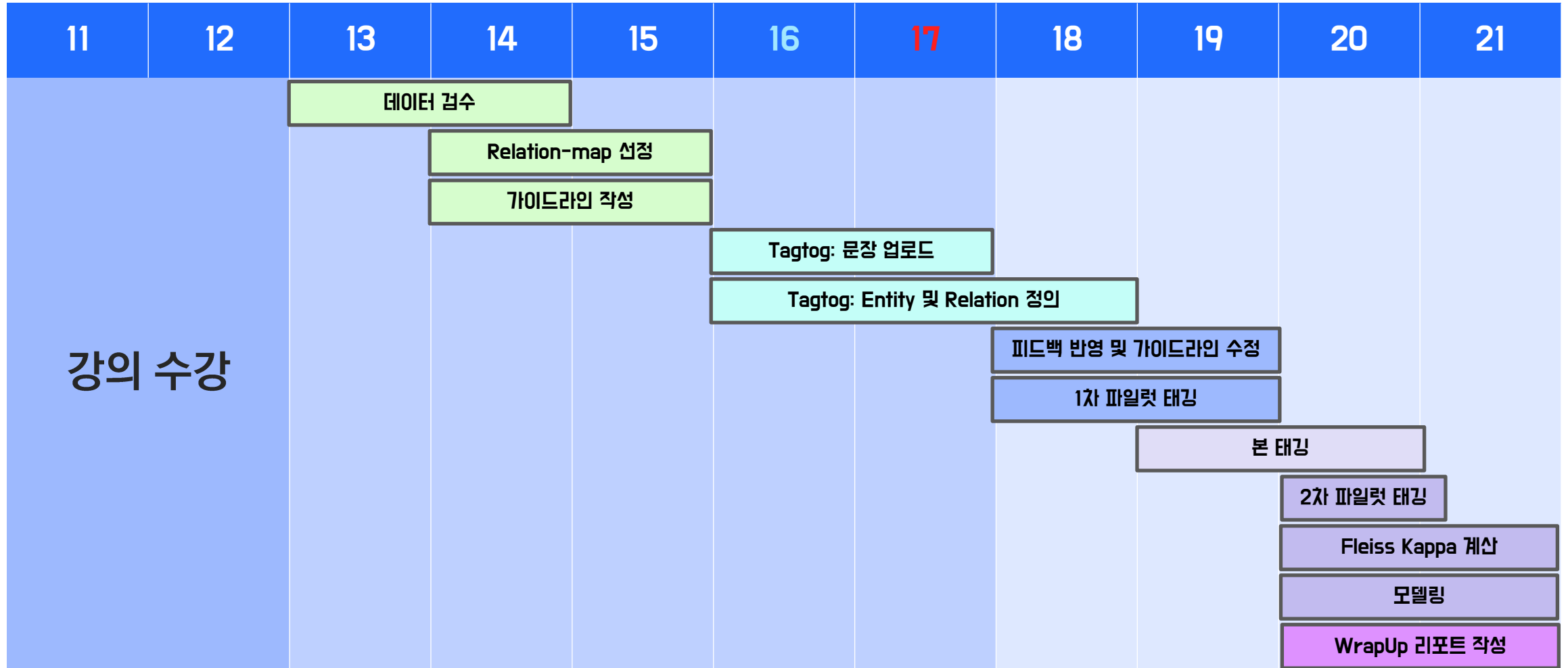


‘섬세킹’ 발표

NLP 9조_MnM (Multi NLP Modal)



데이터 제작 타임라인



데이터 개요

데이터 셋 소개 (러시아 – 우크라이나 전쟁)

2022 러시아의 침공으로 발발한 ‘러시아-우크라이나’ 전쟁을 중심으로 러시아와 우크라이나 간 역사와 외교적 관계를 포괄한 데이터

■ 돈바스 전쟁	■ 루간스크 인민공화국	■ 수미 전투
■ 소련의 붕괴	■ 도네츠크 인민공화국	■ 러시아-우크라이나 전쟁
■ 블라디미르 푸틴	■ 독립국가연합	■ 헤르손 전투
■ 러시아-우크라이나 관계	■ 블로디미르 젤렌스키	■ 2014년 우크라이나 친러시아 분쟁
■ 유로마이단	■ 우익 섹터	■ 체르니히우 전투
■ 유럽 연합	■ 하르키우 전투 (2022년)	■ 돈바스
■ 2021년~2022년 러시아-우크라이나 위기	■ 민스크 협정	■ 스타로빌스크 전투
■ 크림 위기	■ 러시아의 크림반도 합병	■ 마리우폴 공방전
■ 부다페스트 안전 보장 각서	■ 2014년 2월 우크라이나 교통 봉쇄	■ 뱀섬 전투
■ 2022년 러시아의 우크라이나 침공	■ 노드스트림	■ 초르노빌 전투
■ 우크라이나	■ 2014년 우크라이나 혁명	■ 키아우 공방전 (2022년)
■ 북대서양 조약 기구	■ 호스토멜 공항 전투	■ 키아우
■ 핵확산방지조약	■ 엘리토폴 전투	
■ 러시아	■ 미콜라이우 전투	

데이터 없음

선정 이유

annotated된 데이터를 통해 실제 NLP Task에 적용시킬 수 있을 것 같은 주제라고 생각

<35개 텍스트 – 약 2100개 문장>

개선 기회

기존 KLUE 에서의 Relation 관계로는 인사이트를 도출할 만한 Entity와 Relation 선택에 제약

방향 설정

해당 데이터를 통해 알고자 하는, 또 알 수 있는 **정보**와 **관계**가 무엇일 지에 대해서 정의 필요

Relation-Map

Relation-map 선정

<코퍼스 전수 조사>

돈바스 전쟁.txt	2022년 3월 23일 오후 5:53	98KB
소련의 붕괴.txt	2022년 3월 23일 오후 5:53	92KB
볼로디미르 푸틴.txt	2022년 3월 23일 오후 5:53	26KB
러시아-우크라이나 관계.txt	2022년 3월 23일 오후 5:53	13KB
유로마이단.txt	2022년 3월 23일 오후 5:53	10KB
유럽 연합.txt	2022년 3월 23일 오후 5:53	10KB
2021년~2022년...크라이나 위기.txt	2022년 3월 23일 오후 5:53	9KB
크림 위기.txt	2022년 3월 23일 오후 5:53	9KB
부다페스트 안전 보장 각서.txt	2022년 3월 23일 오후 5:53	8KB
2022년 러시아의...크라이나 침공.txt	2022년 3월 23일 오후 5:53	8KB
우크라이나.txt	2022년 3월 23일 오후 5:53	5KB
북대서양 조약 기구.txt	2022년 3월 23일 오후 5:53	5KB
핵확산방지조약.txt	2022년 3월 23일 오후 5:53	7KB
러시아.txt	2022년 3월 23일 오후 5:53	6KB
루간스크 인민공화국.txt	2022년 3월 23일 오후 5:53	6KB
도네츠크 인민공화국.txt	2022년 3월 23일 오후 5:53	4KB
독립국가연합.txt	2022년 3월 23일 오후 5:53	3KB
볼로디미르 젤렌스키.txt	2022년 3월 23일 오후 5:53	3KB
우익 섹터.txt	2022년 3월 23일 오후 5:53	3KB
하르키우 전투 (2022년).txt	2022년 3월 23일 오후 5:53	3KB
민스크 협정.txt	2022년 3월 23일 오후 5:53	3KB
러시아의 크림반도 합병.txt	2022년 3월 23일 오후 5:53	3KB
2014년 2월 우크라이나 교토 봉쇄.txt	2022년 3월 23일 오후 5:53	2KB
노드스트림.txt	2022년 3월 23일 오후 5:53	2KB
2014년 우크라이나 혁명.txt	2022년 3월 23일 오후 5:53	2KB
후스토멜 교차 전투.txt	2022년 3월 23일 오후 5:53	2KB
멜리토폴 전투.txt	2022년 3월 23일 오후 5:53	2KB
미콜라이우 전투.txt	2022년 3월 23일 오후 5:53	2KB
수미 전투.txt	2022년 3월 23일 오후 5:53	1KB
러시아-우크라이나 전쟁.txt	2022년 3월 23일 오후 5:53	1KB
헤르손 전투.txt	2022년 3월 23일 오후 5:53	1KB
2014년 우크라이나...친러시아 분쟁.txt	2022년 3월 23일 오후 5:53	1KB
체르니히우 전투.txt	2022년 3월 23일 오후 5:53	1KB
돈바스.txt	2022년 3월 23일 오후 5:53	675바이트
스타로빌스크 전투.txt	2022년 3월 23일 오후 5:53	465바이트

<팀원 별 Relation 선정>

찬국			
1	관계 없음	no_relation	(*, *)
2	기관: 적대국	org: hostile	(ORG, ORG)
10			
재학	전체 : 379	no_relation : 296	title: 38
1	관계 없음	no_relation	(*, *)
2	인물: 직업/직함	per: title	(PER, POH)
3	인물: 별칭	per: alternate_names	(PER, PER)
10	인물: 동료	per: colleagues	(PER, PER)
태일			
1	단체: 별칭	org: alteranate_names	(ORG, ORG)
2	관계 없음	No Relation	(*, *)
10			
성진			
1	인물: 법령, 조약	per: Act(법령, 조약)	(PER, ACT)
2	법령, 조약: 연도	법령, 조약 - Year(연도)	(ACT, DAT)
6	단체: 본부 도시	org: city_of_headquarter	(ORG, LOC)
7			(?, NOH)
나연			
1	단체: 시작 날짜	org: date_of_start	(ORG, DAT)
21	단체: 상위 단체	org: member_of	(ORG, ORG)

<Relation 정리 및 정의>

	class_name (ko)	class_name (en)	direction (sub, obj)	description
1	관계 없음	no_relation	(*, *)	관계가 없거나, 주어진 관계로 정의할 수 없음
2	인물: 직업/직함	per: title	(PER, POH)	Object는 Subject의 직업/직함
3	단체: 상위 단체	org: member_of	(ORG, ORG)	Object는 Subject가 속한 단체
4	단체: 별칭	org: alternate_names	(ORG, ORG)	Object는 Subject의 또 다른 이름
5	단체: 구성원	org: top_members/employees	(ORG, PER)	object는 subject에 소속된 구성원
6	단체: 적대 관계	org: hostile	(ORG, ORG)	Object는 Subject와 적대관계
7	단체: 우호 관계	org: friendly	(ORG, ORG)	Object는 Subject와 우호관계
8	단체: 자산	org: property	(ORG, [ORG/PER/POH])	Object는 Subject의 자산
9	사건: 장소	eve: place	(EVE, LOC)	Object는 Subject가 발생한 장소
10	사건: 날짜	eve: date	(EVE, DAT)	Object는 Subject의 발생 날짜
1 전쟁이라는 역사를 다루고 있기 때문에 인물보다는 국가에 대한 Entity가 주로 등장함. 2 전쟁이라는 텍스트의 특성 상 '적대 관계', '우호 관계'의 나라나 기관이 주요 Entity로 등장 3 국가의 병력, 자원, 시설, 무기 등의 '자산'이 주요 Entity로 등장 4 역사적 흐름의 텍스트가 주가 되어 특정 사건을 기준으로 '날짜'나 '장소'가 주요 Entity로 등장				

가이드라인 작성

- Relation-map을 토대로 가이드라인 작성
- 관계에 대한 구체적 지침은 Relation-map **회의 내용** 토대로 '자주 묻는 질문' 작성

'러시아-우크라이나 전쟁' 분야 관계 추출 태스크 가이드라인

본 가이드라인은 '러시아-우크라이나 전쟁' 분야의 관계 추출(Relation Extraction) 태스크를 위한 데이터 제작 과정을 설명하기 위해 작성되었습니다. [Annotation에 참여하기 전에 반드시](#) 숙지해주시기 바랍니다.

1. '러시아-우크라이나 전쟁' 분야의 관계 추출 태스크 알아보기

2. Annotation 가이드라인

2-1) 관계의 방향성

2-2) 관계 없음

2-3) 외부 지식의 활용

2-4) 관계의 시점

3. Annotation 환경

3-1) 작업자 환경 준비

3-2) 관계 클래스 선택

3-3) Data Error (자동으로 entity를 잡는 프로세스에서는 필수적이지만, manual하게 잡을 경우에는 option)

3-4) Hate / Bias / Privacy (중요한 문제이지만 위키피디아 문서에서는 두드러지지 않을 것)

4. 자주 묻는 질문

1. '러시아-우크라이나 전쟁' 분야의 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 10개의 클래스 중 하나로 분류합니다. 관계 클래스는 Text의 특성을 반영하여 1) "단체 중심 관계" 6개 2) "사건 중심 관계" 2개 3) "인물 중심 관계" 1개 4) "관계_없음" 1개로 구성됩니다.

- 1) 단체 중심 관계는 Subject Entity가 국가, 기관, 정당 등의 단체일 때 발생할 수 있는 관계로서 "단체:구성원", "단체:별칭", "단체:적대_관계", "단체:우호_관계", "단체:상위_단체", "단체:자신" 이 이에 해당합니다.
- 2) 사건 중심 관계는 Subject Entity가 특정 지을 수 있는 전투, 회담, 조약, 행사 등의 사건일 때 발생할 수 있는 관계로서 "사건:장소", "사건:날짜" 가 이에 해당합니다. 단순한 명사 그 자체의 전투, 회담, 조약, 사퇴, 임명, 취임 등은 사건으로 볼 수 없습니다.
- 3) 인물 중심 관계는 Subject Entity가 인물일 때 발생할 수 있는 관계로서 "인물:직업/직함"이 이에 해당합니다.
- 4) "관계_없음"은 주어진 문장에서 Entity 쌍이 아무 관계가 없음을 의미합니다. (2-2) 관계 없음 에서 자세히 설명)
- 전체 10개의 관계 클래스에 대한 자세한 설명 및 예시는 [링크](#)를 꼭 참고해주세요!

Example #1:

Object Subject
미국 대통령 조 바이든과 러시아 대통령 블라디미르 푸틴이 전화 통화로 존재시키려 했지만 성과 없이 협상에 실패하였다.

위의 문장에서 Entity 쌍을 이루는 Subject Entity는 "블라디미르 푸틴"이며, Object Entity는 "대통령"입니다. 또한 위 문장에서 "블라디미르 푸틴"과 "대통령"사이의 의미적 관계는 "인물:직업/직함" 클래스라고 분류할 수 있습니다.

2. Annotation 가이드라인

4. 자주 묻는 질문

Q1. "인물:직업/직함"의 범위는 어디까지인가요?

- 2-4)관계의 시점과 같이 주변 문맥이나 배경, 현 상황을 제외하고 주어진 문장을 기준으로 관계를 구분합니다.
- "푸틴은 1990년 5월에 레닌그라드 시장을 역임하고 있던 아나톨리 슐차코로부터 국제 관계 담당 고문으로 임명되었다."와 "푸틴은 1994년 3월에 상트페테르부르크 연방특별시장 제1부위원장으로 임명되었다."처럼 시간의 흐름에 따라 직업/직함이 달라질 수 있습니다.

Q2. "단체:상위_단체"의 범위는 어디까지인가요?

- 국가, 도시, 기관 등 '단체' 간 소속 관계가 형성 이된다면, "단체:상위_단체"로 구분될 수 있습니다.
- 국가, 도시, 기관의 '단체'와 '인물' 간 관계라면 "단체:구성원"이 맞습니다.
- 예외적으로, 특정 국가의 국민은 행정부에 실제로 근무하고 있지않아, 행정부나 정부의 소속으로 판단하지 않습니다

Q3. "단체:별칭"의 범위는 어디까지인가요?

- 기관, 국가, 의회, 직위 등이 이에 해당할 수 있습니다.
- "시위 도중 마이단이라는 단어가 국민 정치 행동 그 자체를 의미하는 단어로 쓰여지게 되었다" 단어가 다른 의미를 가지게 되었을 경우 이는 "관계_없음"관계입니다. (POH, POH)

Q4. "단체:우호_관계", "단체:적대_관계"는 어떻게 구분하나요?

- 2-4)관계의 시점과 같이 주변 문맥이나 배경, 현 상황을 제외하고 주어진 문장을 기준으로 관계를 구분합니다.
- "공격, 침공, 폭파, 테러" 등 긍정적인 관계를 내포하는 단어가 나왔을 때 "단체:적대_관계"로 정의합니다.
- "지지, 협약, 지원, 교류" 등 긍정적인 관계를 내포하는 단어가 나왔을 때 "단체:적대_관계"로 정의합니다.
- "소련이 붕괴되고, 우크라이나는 러시아로부터 에너지 자원을 공급 받았고, 우크라이나는 가스를 공급받으면서 밀렸던 채무들을 투르레프 Tu-160 등의 핵무기를 러시아로 이전하여 해결했다." 해당 문장에서 우크라이나와 러시아는 "단체:우호_관계"로 볼 수 있습니다.
- "러시아 전차가 우크라이나 국경을 넘었다."와 같은 예시는 공격하려고 국경을 넘은건지, 도우려고 국경을 넘은건지 외부 지식을 사용하지 않는 한 알 수 없기 때문에 "단체:적대_관계"가 될 수 없습니다.

TagTog Entity Relation 설정

- Tagtog 특성 상 Entity와 Relation 관계 설정 시 문제 예상 (Relation에 따른 Entity 개수 증가)
- **SUB Entity**에만 Relation 정의하고 **OBJ Entity** 중복 사용으로 전체 Entity 개수 감소
(엔티티 색상 구분 SUB: 원색 계열 / OBJ: 파스텔 계열)

<Relation_Type (20개)>

Relation Types

Find below the relation types you can use to link your annotated entities.

SUB_PER_NoRe ↔ OBJ_PER

SUB_PER_NoRe

OBJ_PER

관계가 없거나, 주어진 관계로 정의할 수 없음 (사람-사람)

SUB_PER_NoRe ↔ OBJ_ORG

SUB_PER_NoRe

OBJ_ORG

관계가 없거나, 주어진 관계로 정의할 수 없음 (사람-기관)

SUB_PER_NoRe ↔ OBJ_POH

SUB_PER_NoRe

OBJ_POH

관계가 없거나, 주어진 관계로 정의할 수 없음 (사람-사물)

SUB_ORG_NoRe ↔ OBJ_PER

SUB_ORG_NoRe

OBJ_PER

관계가 없거나, 주어진 관계로 정의할 수 없음 (기관-사람)

SUB_ORG_NoRe ↔ OBJ_ORG

SUB_ORG_NoRe

OBJ_ORG

관계가 없거나, 주어진 관계로 정의할 수 없음 (기관-기관)

SUB_ORG_NoRe ↔ OBJ_POH

SUB_ORG_NoRe

OBJ_POH

관계가 없거나, 주어진 관계로 정의할 수 없음 (기관-사물)

<SUB_Entity (12개)>

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

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

SUB_PER_NoRe Subject No_relation Person Delete Save ✓

SUB_ORG_NoRe Subject No_relation Organization Delete Save ✓

SUB_EVE_NoRe Subject No_relation Event Delete Save ✓

SUB_PER_Title Subject Title Person Delete Save ✓

SUB_ORG_MemberOf Subject Member of Organization Delete Save ✓

SUB_ORG_AlterName Subject AlternativeName Organization Delete Save ✓

SUB_ORG_Hostile Subject Hostile Organization Delete Save ✓

SUB_ORG_Topm1Emp Subject TopMember/Employee Organization Delete Save ✓

SUB_ORG_Friendly Subject Friendly Organization Delete Save ✓

SUB_ORG_Property Subject Property Organization Delete Save ✓

SUB_EVE_Place Subject Place Event Delete Save ✓

SUB_EVE_Date Subject Date Event Delete Save ✓

<OBJ_Entity (5개)>

OBJ_PER Object Person Delete Save ✓

OBJ_ORG Object Organization Delete Save ✓

OBJ_POH Object title poh Delete Save ✓

OBJ_LOC Object Location Delete Save ✓

OBJ_DAT Object Date Delete Save ✓

<문장 분배 (약 420개씩)>

tagtog

taeil/boost_camp_NLP_9_data_annotation

Settings Documents Metrics

+ Content

pool	1679.txt	소련군이 바쿠를 장악했으나 1월 20일까지
changuk	1678.txt	이 사건은 소련군이 소련 자국 내의 도시 중
jaehak	1677.txt	소련의 국방부 장관 드미트리 야조프는 바쿠
nayeon	1676.txt	그 외 800명 이상이 부상을 입었으며 수백명
seongjin	1675.txt	아들 중 대부분은 민간인이었다.
taeil	1674.txt	이날 밤 이후로 2월 초까지 계속된 소련군과
	1673.txt	1990년 1월 9일 저녁, 중앙 텔레비전 방송국

Tagging

1차 파일럿 태깅 & 가이드라인 수정

Index	Sentence	Relataion	찬국	재학	나연	성진	태일
태일	1월 23일인 우크라이나 그리스 가톨릭교회가 1946년 소비에트에게 탄압당한 뒤 처음으로 시노드 (종교 회의)를 열었다.		사건: 날짜	사건: 날짜	사건: 날짜	사건: 날짜	사건: 날짜
	8월 30일, 바이조프 제1서기 가 모스크바로 송환되었고 이후 당내 자유투표를 통해 다음 제1서기에 아야즈 위탈리토프가 등극하였다.		인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함
	북대서양 조약기구 또한 크림 공화국과 일대에 긴장감이 높아지고 있음을 지적하며 러시아에 대한 비판 수위를 높였고 유럽위원회 는 우크라이나 내부의 결속력을 다지는 데 모든 지원을 아끼지 않았다고 성명을 냈다.	단체: 우호_관계	단체: 우호_관계	관계: 없음	관계: 없음	단체: 우호_관계	단체: 우호_관계
찬국	유엔 안전 보장 이사회의 13개 이사국은 크림 반도에서 실시된 주민투표가 무효임을 선언한 결의에 찬성하였으나 중화인민공화국은 기권했고, 러시아 가 거부권을 행사하면서 부결되었다.	관계: 없음	단체: 상위_단체	관계: 없음	단체: 적대_관계	관계: 없음	관계: 없음
	우크라이나 에 군사 개입과 크림 반도 의 병합.	관계: 없음	관계: 없음	관계: 없음	관계: 없음	단체: 우호_관계	관계: 없음
	체첸 공화국은 선거 이후 안정되었다.		미태깅	미태깅	미태깅	미태깅	미태깅
재학	소련군 이 바쿠 를 장악했으나 1월 20일까지 사실상 아제르바이잔이 무정부 상태였다.	단체: 적대_관계	단체: 적대_관계	관계: 없음	단체: 적대_관계	관계: 없음	단체: 적대_관계
	이 사건은 소련군 이 소련 자국 내의 도시 중 하나를 강제로 점령한 최초의 군사 행동이었다.	단체: 자산	단체: 상위_단체	단체: 자산	단체: 구성원	관계: 없음 (소련군)	단체: 자산
	3월 11일, 리투아니아 SSR에서 새로 구성된 최고회의 의장 (대통령)으로 사유디스의 지도자인 비타우타스 란즈베르기스 가 선출됨과 동시에 그가 리투아니아 국가 재건 법률을 선포하며 소련 내 소비에트 공화국 중에선 처음으로 소련을 탈퇴하고 독립국을 선언하였다.	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	관계: 없음
성진	우크라이나 내무부 장관인 아르센 아바코프 는 "우리 정보는 따르면, 아침에 스킨드에서 장갑차, 대포, AMV, 전자 및 기타 중장비가 러시아에서 우크라이나 국경을 넘은 것이 확인되었다"라고 말했다.		인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함
	이 전투로 인해 도네츠크 인민 공화국의 부통령인 맥심 페트루킨이 사망하고 대통령 데니스 푸실린 은 부상을 입었다.		인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함	인물: 직업/직함
	정부군 은 낙하산병 과 폭격을 동원하여 반군에 대한 공격을 시작했다.	단체: 구성원	단체: 자산	단체: 구성원	단체: 구성원	단체: 구성원	단체: 구성원
나연	또한, 우크라이나군 에 의한 구타, 납치, 지역 주민들 의 처형에 관한 여러 제보도 있었다.	관계: 없음	관계: 없음	단체: 적대_관계	관계: 없음	관계: 없음	단체: 적대_관계
	반정부군 은 누구도 그들을 도와주거나 지지하지 말라는 메시지를 보내는 목적으로 구타 와 납치를 일상화한다."	태깅X	관계: 없음	관계: 없음	관계: 없음	관계: 없음	관계: 없음
	우크라이나 육군 은 우크라이나 내의 주요 군사세력이며 도네츠크 인민 공화국과 루한스크 인민 공화국 에 맞싸 싸우는 선도적인 군사 중 하나이다.		단체: 적대_관계	단체: 적대_관계	단체: 적대_관계	단체: 적대_관계	단체: 적대_관계



태깅 작업하면서 추가한 FAQ

Q8. "관계_없음" 문장의 경우 어디까지 포함하는가?

- 최대한 entity 두 개와 relation을 찾아서 "관계_없음" 태깅합니다. 하지만 한 문장에 entity가 2개 나오지 않거나 두 entity간에 연관이 아예 없는 경우에는 아예 태깅을 하지 않고 넘어갑니다.

Q9. "인물:직업/직함"과 "단체:구성원" 간 유사성이 있어보이는데 어떤 기준으로 태깅하나요?

- 미국 대통령 조 바이든과 러시아 **대통령 블라디미르 푸틴**이 전화 통화로 증재시키려 했지만 성과 없이 협상에 실패하였다.
- object entity에는 문장에서 나타나는 핵심 관계에 따라 태깅합니다. '블라디미르 푸틴'의 직위(대통령)가 중요한 관계라고 생각한다면 "인물:직업/직함"으로, '블라디미르 푸틴'의 소속이 중요한 관계라고 생각한다면 "단체:구성원"으로 태깅합니다.

Q10. entity에 조사가 들어가는 경우 어디까지 entity에 해당하는가?

- Entity에 조사(관형격조사 '의')가 들어가는 경우 관형사 + 명사로 취급합니다. 따라서 '러시아의 연방 정부'의 경우 "연방 정부"를 하나의 Entity로 보고, '러시아 연방 정부'의 경우 그 자체를 하나의 Entity로 볼 수 있습니다.

Q11. entity 태깅에 복수형도 포함하나요?

- 러시아 국방부는 이 사건을 두고 **병사들**이 실수로 국경을 넘어서 일어난 일이라고 발표했다.
- '병사' 대신 '병사들'로 복수형 표기로 entity 태깅을 진행합니다.

Q12. '후신'(이름이 변경되는 것)된 기관명도 "단체:별칭"으로 볼 수 있나?

- 후신된 경우 기관의 이름과 수행이 달라지는 것으로 보며, 같은 시간대에 서로 다른 이름으로 불려지는게 아니기 때문에 "단체:별칭"으로 보지 않습니다.

<파일럿 태깅>

본 태깅 작업 (1770개 [=2100 - 330])

헤르손 전투(,)는 2022년 2월 24일 러시아의 우크라이나 침공 당시 벌어진 전투이다.

1991년 6월에는 **상트페테르부르크 시청 산하 대외 관계 위원회** 위원장으로 임명되면서 상트페테르부르크의 대외 관계 향상, 외국인 투자 촉진 사업에 참여했다.

2008년 5월 7일 **드미트리 메드베데프** 러시아 정부 제1부총리가 러시아 제3대 대통령으로 취임하자, 러시아 총리로 지명받아 2008년 5월 8일에 러시아 역사상 최초로 전임 러시아 연방 대통령 출신 현직 러시아 총리(러시아 총리직은 2번째)가 되었다.

<가이드라인 수정>

Fleiss Kappa 및 모델링

2차 파일럿 태깅 및 Fleiss Kappa 계산

Index	sentence	subject_entity	object_entity	찬국	재학	나연	태일	성진
1	이듬해 <OBJ 3월 26일> 열린 <SUB 정식 러시아 연방 대통령선거>에서 러시아 연방의 2번째 대통령으로 당선되었다.'	EVE	DAT	사건: 날짜 ▾	사건: 날짜 ▾	사건: 날짜 ▾	사건: 날짜 ▾	사건: 날짜 ▾
2	1922년, <SUB 우크라이나>는 <OBJ 소비에트 사회주의 공화국 연방>의 창립 국가였으며, 러시아 제국이 무너져 결국은 우크라이나어 사용 금지도 철폐되었다.'	ORG	ORG	단체: 상위_단 ▾	관계_없음 ▾	단체: 상위_단체 ▾	관계_없음 ▾	단체: 상위_단체 ▾
3	또한 <SUB 돈바스 전쟁>에서는 <OBJ 우크라이나> 정부군을 지지했다.'	EVE	ORG	관계_없음 ▾	관계_없음 ▾	관계_없음 ▾	관계_없음 ▾	관계_없음 ▾
200	또한 <OBJ 유럽>은 아니지만 몇몇 유럽 연합 회원국의 속령도 <SUB 유럽 연합>의 일부이다.'	ORG	ORG	관계_없음 ▾	단체: 우호_곤 ▾	관계_없음 ▾	관계_없음 ▾	관계_없음 ▾

<2차 파일럿 태깅 (200개)>

	# raters	# categories	# subjects	PA	PE	Fleiss' Kappa
1차 파일럿 태깅	5	10	100	0.7570	0.1546	0.713
2차 파일럿 태깅	5	10	200	0.8059	0.2243	0.75

<1, 2차 파일럿 태깅 Fleiss' Kappa>

모델링

- 전체 1770개 데이터를 9:1 비율로 Stratified Split을 사용해 train/ test 데이터로 나눔
- 이후 모델 학습에서 train data를 다시 9:1 비율로 나누어 각각 train / validation으로 사용
- 이전 KLUE RE Task의 모델을 활용해, Label만 변경 후 모델링 진행

micro-f1 score

valid 87.097
test 88.152

<자체 평가 의견>

김태일_T3063	완전히 배경 지식이 없는 분야여서 최대한 편견 없이 관계들을 정의할 수 있어 데이터 제작 프로세스에 관한 학습이 잘 되었다. 하지만 일부 특정 서브젝트 오브젝트 타입의 조합이면 거의 무조건적으로 특정 관계가 나오는 태깅들도 있어서 다음 데이터 제작에 참여하게 된다면 더 다양하게 태깅하면 좋을 것 같다.
문찬국_T3076	처음하는 데이터 제작 프로젝트였지만, 꽤 수월하게 진행했던 것 같다. '모두가 만족하는 가이드라인은 없다'든가 초기 가이드라인의 중요성에 대해서 알 수 있었던 것 같다. 개인적으로 엔티티 간 관계에만 신경을 쓰다보니, 엔티티 선정 기준을 제대로 정하지 못했던 부분이 조금 아쉬웠다. 다음 번에 데이터 제작을 하게 된다면, 그때는 초기에 데이터 제작 방향을 조금 더 확실히 하는데 시간을 더 써야겠다.
이재학_T3161	주제를 선택한 근거가 부족했던 부분이 아쉬웠다. 하지만 작업을 진행하며 '전쟁'이라는 주제로 인해 'ORG', 'PER' 등의 Entity Type이 자주 등장하여 원활한 데이터셋을 구축할 수 있었으로써 주제 선택에 대한 근거를 찾을 수 있었고, 이전의 RE 대회에서 데이터들의 Mislabeling 및 유사한 데이터들이 나오는 이유를 간접적으로 찾을 수 있었다. 배경지식없이 문장만 보며 태깅을 해야 하는 부분이 어려웠지만, 팀원들과 기준을 잡아가며 IAA를 높여가는 과정에서 데이터 제작과정을 이해할 수 있어서 좋았다.
하성진_T3230	현재 당장 시급하면서도, 복잡한 국제 정세 및 관계가 내포된 주제를 기반으로 RE 데이터를 구축하는 것이 흥미로웠다. 이 주제 속에서 다양한 관계 추출을 위해서 새로운 엔티티와 릴레이션 정의를 시도해보는 것은 좋았지만, 레퍼런스가 없다보니 엔티티 정의후 시행착오가 많았으며, 엔티티 관계에 내포된 중요한 관계 정보를 추출했다고 평가하기에는 아쉬움이 남는다. 이번 경험을 계기로, 좀 더 데이터의 목적성을 명확히하고 가이드라인을 확립하는데 좀 더 많은 시간을 사용할 수 있기를 기대해본다.
한나연_T3250	데이터 제작을 단순히 라벨링 작업이라고 생각했는데 큰 오산이었다. 동일한 데이터에 대해 서로 다른 entity와 relation을 정의할 수 있고 데이터에 다양한 relation이 존재해 합의점을 찾는데 상당한 시간을 쏟았다. 팀원들끼리 왜 해당 관계로 태깅했는지 토의하면서 주제에 더 관련된 데이터를 구축할 수 있는 방향으로 태깅을 진행했다. 벤치마크 데이터셋을 가져와 사용하거나 조금 변경하는 정도가 전부였는데, 이번 데이터 제작 대회를 통해 전반적인 데이터셋 제작 과정을 알게 되었고, 오픈 데이터셋의 소중함을 더 느꼈다.

<결론>

- 이번 프로젝트에서 전반적인 데이터셋 제작 과정을 경험해볼 수 있었다.
- 초기 데이터 제작의 목적성과 가이드라인 확립에 더 많은 시간 투자 필요

Thanks To



김태일



문찬국



이재학



하성진



한나연

QnA