

TTA Standard

정보통신단체표준(국문표준)

TTAx.xx-xx.xxxx/R1

제정일: 2015년 xx월 xx일

(1차 오류정정: 20xx년 xx월 xx일)

(2차 오류정정: 20xx년 xx월 xx일)

개체명 태그세트 및 태깅 말뭉치

Tag Set and Tagged Corpus for Named
Entity Recognition



한국정보통신기술협회
Telecommunications Technology Association

개체명 태그세트 및 태깅 말뭉치

Tag Set and Tagged Corpus for Named Entity Recognition



본 문서에 대한 저작권은 TTA 에 있으며, TTA 와 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

Copyright© Telecommunications Technology Association 2015. All Rights Reserved.

서 문

1. 표준의 목적

대규모 말뭉치(corpus)는 자연어 처리 분야에서 가장 기초적인 언어 자원으로 그 중요성이 높아지고 있다. 특히 말뭉치를 분석하여 가공한 개체명 태깅 말뭉치(tagged corpus)는 개체명 인식기 개발에 있어서 필수적인 언어 자원으로 그 구축과 관리에 많은 비용이 소요되므로 연구 개발자 간에 공유가 가능해야 하고, 기존에 구축되어 있는 것을 재사용할 수 있어야 한다. 본 표준은 개체명 태깅 말뭉치의 재사용 및 공유를 가능하게 하는 표준 태그세트 및 개체명 태깅 지침을 제공하기 위하여 작성되었다.

2. 주요 내용 요약

개체명 인식은 인명, 지명, 기관명 등과 같은 고유명사나 명사구를 의미한다. 본 권고안은 각 개체명의 태그세트에 대한 정의와 분류기준을 제공한다. 또한 개체명 태깅 말뭉치를 구축하면서 발생할 수 있는 다양한 애매성에 대해서 명확한 기준을 제시해서 태깅 말뭉치 구축 시에 발생할 수 있는 문제점을 해소하는 것에 목적이 있다.

본 권고안에서는 정보검색이나 질의응답 등의 다양한 자연어 처리 분야에서 활용 가능한 개체명 태그세트를 제시한다. 표준 개체명 태그세트는 다양한 응용분야에서 사용할 수 있도록 인명, 지명, 기관명 등의 대분류 보다 세분화된 태그세트로 구성되어 있다.

3. 표준 적용 산업 분야 및 산업에 미치는 영향

언어분석은 정보 추출, 정보 검색, 질의 응답 등의 자연어 처리 분야에서 기반이 되는 기술이며, 본 표준은 언어분석 과정 중 개체명 인식에 대한 것이다. 정보 추출, 질의응답 등의 응용분야 기술은 개체명 인식의 태그세트에 매우 의존적이므로 응용 분야의 필요에 부합되도록 개체명 태그세트를 확장/정제하고 품사 태깅 말뭉치 작성용 표준안을 마련할 필요가 있다. 응용 분야의 필요에 적합한 품사 태그세트와 말뭉치 구축 표준안이 나오면, 표준안에 따르는 개체명 태깅 말뭉치가 다양한 기관에서 만들어지고 공유되면서 개체명 인식기 개발 생태계가 자연스럽게 조성되고, 나아가서는 개체명 인식기를 활용하는 응용 분야의 기술 발전에도 기여할 것이다.

4. 참조 표준(권고)

4.1. 국외 표준(권고)

- 해당 사항 없음

4.2. 국내 표준

- 해당 사항 없음

5. 참조 표준(권고)과의 비교

5.1. 참조 표준(권고)과의 관련성

- 해당 사항 없음

5.2. 참조한 표준(권고)과 본 표준의 비교표

- 해당 사항 없음

6. 지식재산권 관련사항

본 표준의 '지식재산권 요약서' 제출 현황은 TTA 웹사이트에서 확인할 수 있다.

※본 표준을 이용하는 자는 이용함에 있어 지식재산권이 포함되어 있을 수 있으므로, 확인 후 이용한다.

※본 표준과 관련하여 접수된 요약서 이외에도 지식재산권이 존재할 수 있다.

7. 시험인증 관련사항

7.1. 시험인증 대상 여부

- 해당 사항 없음

7.2. 시험표준 제정 현황

- 해당 사항 없음

8. 표준의 이력 정보

8.1. 표준의 이력

판수	제정.개정일	제정.개정내역
제 1 판		제정 TTAx.xx-xx.xxxx

8.2. 주요 개정 사항

- 해당 사항 없음

Preface

1. Purpose of Standard

One of the most essential language resources, large-scale corpora becomes more important in the Natural Language Processing research and application area. Especially, tagged corpora obtained by attaching tags to the named entities in the raw corpora are basic language resources for named entity recognition and require tremendous cost for their development and management. As a result, it is desirable that tagged corpora should be reusable and sharable among the research community. The purpose of this standard is to provide commonly usable guidelines for the Named Entity Tag Set for the reuse and sharing of tagged corpus.

2. Summary of Contents

Named entities are phrases that contain the names of persons, organizations, locations, times and quantities. The tagging of Named Entities, the names of particular things or classes, is regarded as an important component technology for many NLP applications. These applications include question answering (QA), summarization, information retrieval (IR) and information extraction (IE). This standard is to provide commonly usable guidelines for the Named Entity Tag Set and tagged corpus.

3. Applicable fields of industry and its effect

Language analysis is the essential technology in the Natural language processing areas for information retrieval, information search and Question Answering. The standard is related with the technology of named entity recognition, the core technology among various language analysis technologies. Since the application technologies such as

information retrieval are heavily dependent on the named entity tag sets, it is required to extend and refine the named entity tag set and to establish a standard to construct tagged corpus satisfying the needs of the applications. The standardized named entity tag set and tagged corpus guidelines will contribute to the development of the ecosystem of the named entity recognition and technology enhancement of various application areas by producing and sharing named entity tagged corpus in many different institutes.

4. Reference Standards (Recommendations)

4.1. International Standards (Recommendations)

- None

4.2. Domestic Standards

- None

5. Relationship to Reference Standards(Recommendations)

5.1. Relationship of Reference Standards

- None

5.2. Differences between Reference Standard(recommendation) and this Standard

- None

6. Statement of Intellectual Property Rights

IPRs related to the present document may have been declared to TTA. The information pertaining to these IPRs, if any, is available on the TTA Website.

No guarantee can be given as to the existence of other IPRs not referenced on the TTA website.

And, please make sure to check before applying the standard.

7. Statement of Testing and Certification

7.1. Object of Testing and Certification

– None

7.2. Standards of Testing and Certification

– None

8. History of Standard

8.1. Change History

Edition	Issued date	Outline
The 1st edition		Established TTAx.xx-xx.xxxx

8.2. Revisions

– None

목 차

1. 개 요	1
2. 표준의 구성 및 범위	1
3. 참조 표준(권고)	2
4. 용어정의	2
5. 태그세트 분류 체계	3
6. 태깅 말뭉치 구축 지침	8
6.1. 인물	8
6.2. 학문분야	8
6.3. 이론	12
6.4. 인공물	15
6.5. 기관	20
6.6. 지역	26
6.7. 문명	30
6.8. 날짜	36
6.9. 시간	38
6.10. 수량	39
6.11. 이벤트	44
6.12. 동물	46

6.13. 식물	48
6.14. 물질	50
6.15. 용어	52
7. 기타 태깅 원칙	56

Contents

1. Introduction	1
2. Constitution and Scope	1
3. Reference Standards (Recommendations)	2
4. Terms and Definitions	2
5. Tag-set classification criteria	3
6. Tagging guideline	8
6.1. PERSON	8
6.2. STUDY_FIELD	8
6.3. THEORY	12
6.4. ARTIFACTS	15
6.5. ORGANIZATION	20
6.6. LOCATION	26
6.7. CIVILIZATION	30
6.8. DATE	36
6.9. TIME	38
6.10. QUANTITY	39
6.11. EVENT	44
6.12. ANIMAL	46

6.13. PLANT	48
6.14. MATERIAL	50
6.15. TERM	52
7. Other information	55

개체명 태그세트 및 태깅 말뭉치 표준

(Tag Set and Tagged Corpus for Named Entity Recognition)

1. 개요

본 문서는 세부분류 개체명 태그세트 표준과 태깅 말뭉치 구축 지침서이다. 본 표준은 개체명 태그세트 표준을 제시할 뿐만 아니라, 각 개체명 태그세트에 대한 태깅 말뭉치 구축에 대한 지침도 제시한다. 따라서, 세부분류 개체명의 명확한 정의, 태깅 정책, 태깅 시 고려될 사항, 예상되는 문제점 등에 대해 기술되어 있다.

본 문서에서 기술하는 세부분류 개체명 범위는 명사, 복합명사 그리고 일부 명사구 형태뿐 아니라 동사구나 절 형태의 개체명이 될 수 있는 모든 형태를 대상으로 한다. 형태소분석은 명사, 동사, 형용사, 부사, 조사, 어미 등의 형태소 단위 및 적합한 품사를 인식하는 것이 목표인 반면, 개체명인식은 “홍길동”(명사)과 같이 1개의 형태소가 개체명 단위가 될 수도 있지만 “10시부터 12시까지”(명사구)와 같이 2개 이상의 형태소로 구성된 형태가 개체명 단위가 될 수도 있다.

세부분류 개체명의 대분류는 인물(PERSON), 학문분야(STUDY_FIELD), 이론(THEORY), 인공물(ARTIFACTS), 기관(ORGANIZATION), 지역(LOCATION), 문명(CIVILIZATION), 날짜(DATE), 시간(TIME), 수량(QUANTITY), 이벤트(EVENT), 동물(ANIMAL), 식물(PLANT), 물질(MATERIAL), 용어(TERM)로 모두 15개로 구성되어 있으며, 이들은 각각 세부 범주로 세분화되어 있다.

개체명 관련 표준으로 2012년 12월 21일에 제정된 “관계형 기술 지식 추출 모델 및 데이터 포맷”(TTAK.KO-10.0642)이 있다. “TTAK.KO-10.0642”의 경우에는 기술 문헌에 출현하는 개체 위주로 13개 대분류로 개체 유형을 정의한 반면, 본 표준은 기술 문헌 이외의 다양한 분야에 등장하는 개체유형을 중심으로 15개 대분류, 146개 세부 개체유형을 정의하고, 각 세부유형 별로 태깅 시에 발생할 수 있는 애매성을 해소할 수 있도록 태깅 지침을 제시한다. “TTAK.KO-10.0642”은 기술 문헌에 출현하는 중요한 기술을 식별하기 위한 표준으로 지식기술 개체 및 기술 개체 간 의미적 연관 관계를 추출하기 위한 자동화 모델을 제시한 반면, 본 표준은 질의응답 시스템 등의 응용분야에서 사용할 수 있는 146개의 세부 개체유형을 제시하는 것에 중점을 두고 있다.

2. 표준의 구성 및 범위

본 표준은 말뭉치를 분석하여 개체명 단위로 분할할 때, 각 개체명에 부착되는 태그세트에 대한 정의와 분류기준을 제공한다.

본 표준에서 제시하는 개체명 태그세트의 가장 주된 용도는 자동 개체명 인식을 위한 학습자료로 활용되는 개체명 태깅 말뭉치 작성용이다. 태깅 말뭉치 구축 시에는 대상 범위가 애매하거나, 대상 개체명이 2개 이상의 태그로 부착이 가능한 상황 등의 애매한 상

황이 많이 발생한다. 본 표준에서는 개체명 태깅 시에 발생할 수 있는 애매한 상황들을 해소할 수 있도록 각 태그 별로 태깅 시에 참고할 수 있는 태깅 지침을 기술한다.

3. 참조 표준(권고)

- 해당 사항 없음

4. 용어정의

4.1. 언어자원(language resource)

최근에 유행하는 경험적이고 관찰적인 자연언어처리 기법에서 귀납적으로 언어지식을 추출하기 위해 필요한 다양한 언어현상을 담고 있는 기초자료이며, 전자사전, 말뭉치, 품사 태깅 말뭉치, 개체명 태깅 말뭉치 등을 말한다. 이 중, 본 표준안과 직접적으로 관련이 있는 ‘말뭉치’, ‘개체명 태깅 말뭉치’에 대해서만 정의를 내리기로 한다.

4.2. 말뭉치(corpus)

신문, 소설, 교과서 등 각 분야의 저서나 출판물을 정해진 저장양식과 표현양식에 따라 전자화(machine-readable)한 텍스트 모음을 말한다.

4.3. 개체명 인식(named entity recognition)

문장을 기본단위로 분석하여, 문장에 출현한 모든 개체명의 경계를 인식하고, 해당 개체명이 어떤 태그로 분류될 수 있는지를 자동으로 인식하고, 개체명 인식 결과를 출력하는 것을 말한다. 예를 들면, “오늘은 한국으로 돌아간다”라는 입력 문장에 대해 개체명인식기는 <오늘:DT_DAY>, <한국:LCP_COUNTRY>을 개체명으로 인식해서 출력할 수 있다.

4.4. 개체명 태그세트(named entity tag set)

개체명 인식 대상으로 분류될 수 있는 개체명 태그의 분류 체계를 말한다.

4.5. 태그(tag)

태그세트를 이루는 구성요소들로서 태깅의 관점에서 분류한 각각의 개체명을 상징하는 암호화된 표식을 말한다.

4.6. 태깅(tagging)

말뭉치의 문장들을 분석해서, 문장에 나타난 개체명의 경계를 인식하고, 주변 문맥을 고려해서 가장 적합한 개체명 태그를 하나만 부여하는 행위를 말한다.

4.7. 개체명 태깅 말뭉치(named-entity tagged corpus)

정해진 개체명 태그세트 및 태깅 원칙에 따라 말뭉치의 문장에 나타난 개체명의 경계를 표시하고, 정해진 표현양식과 할당 기준 및 문맥에 따라 각 개체명에 1 개의 태그를 부착한 형태의 텍스트 모음을 말한다.

5. 태그세트 분류 체계

대분류	세분류	정의
PERSON (PS)	PS_NAME	- 사람 이름
LOCATION (LC)	LC_OTHERS	- LC계열의 세부유형이 아닌 기타 장소
	LCP_COUNTRY	- 국가명
	LCP_PROVINCE	- 도, 주 지역명
	LCP_COUNTY	- 군, 면, 읍, 리, 동 등과 같은 세부 행정구역명
	LCP_CITY	- 도시명
	LCP_CAPITALCITY	- 수도명
	LCG_RIVER	- 강, 호수, 연못
	LCG_OCEAN	- 해양, 바다
	LCG_BAY	- 반도, 만
	LCG_MOUNTAIN	- 산, 산맥, 능선, 고개/재, 봉
	LCG_ISLAND	- 섬, 제도
	LCG_CONTINENT	- 대륙
	LC_TOUR	- 관광명소
	LC_SPACE	- 천체 명칭
ORGANIZATION (OG)	OG_OTHERS	- 기타 기관/단체
	OGG_ECONOMY	- 경제 관련 기관/단체, 기업
	OGG_EDUCATION	- 교육 기관/단체, 교육관련 기관
	OGG_MILITARY	- 군사 기관/단체 및 유형, 국방기관
	OGG_MEDIA	- 미디어 기관/단체, 방송관련 기관/기업
	OGG_SPORTS	- 스포츠 기관/단체
	OGG_ART	- 예술 기관/단체
	OGG_MEDICINE	- 의학/의료 기관/단체
	OGG_RELIGION	- 종교 기관/단체, 종파 포함
	OGG_SCIENCE	- 과학 기관/단체
	OGG_LIBRARY	- 도서관 및 도서관 관련 기관/단체
	OGG_LAW	- 법률 기관/단체
	OGG_POLITICS	- 정부/행정 기관, 공공기관, 정치기관
	OGG_FOOD	- 음식 관련 업체/회사
	OGG_HOTEL	- 숙박 관련 업체
ARTIFACTS (AF)	AF_CULTURAL_ASSET	- 건물/서적/작품 등 문화재
	AF_BUILDING	- 건축물/토목건설물, 운동장 이름, 아파트, 다리, 등대,

		분수
	AF_MUSICAL_INSTRUMENT	- 악기 명칭
	AF_ROAD	- 도로/철로 명칭
	AF_WEAPON	- 무기 명칭
	AF_TRANSPORT	- 교통수단/자동차/선박 모델 및 유형, 운송수단, 놀이기구
	AF_WORKS	- AFW의 세부 작품명에 해당하지 않는 기타 작품명
	AFW_DOCUMENT	- 도서/서적 작품명 - 지리서, 지도 - 의학관련 서적 - 종교서적, 각 종교의 경전 - 철학관련 서적 - 시/소설/희곡 등의 문학 작품명 - 역사서적 - 기타 서적
	AFW_PERFORMANCE	- 춤/무용 작품명 및 춤 종류 - 연극/공연명/가극
	AFW_VIDEO	- 영화 작품명 - TV 프로그램 이름
	AFW_ART_CRAFT	- 미술 작품명(그림/미술품)
	AFW_MUSIC	- 음악 작품명
	AF_WARES	- 상품/제품 이름
DATE (DT)	DT_OTHERS	- DT 계열의 세부유형이 아닌 기타 날짜
	DT_DURATION	- 기간
	DT_DAY	- 날짜/절기
	DT_MONTH	- 달
	DT_YEAR	- 년
	DT_SEASON	- 계절
	DT_GEOAGE	- 지질시대
	DT_DYNASTY	- 왕조시대
TIME (TI)	TI_OTHERS	- 기타 시간
	TI_DURATION	- 기간
	TI_HOUR	- 시각
	TI_MINUTE	- 분
	TI_SECOND	- 초
CIVILIZATION (CV)	CV_NAME	- 문명/문화 명칭
	CV_TRIBE	- 민족/종족 명칭, 국가를 구성하는 국민
	CV_SPORTS	- 스포츠/레포츠/레저 명칭
	CV_SPORTS_INST	- 스포츠 용품/도구

	CV_POLICY	- 제도/정책 명칭
	CV_TAX	- 조세 명칭
	CV_FUNDS	- 연금, 기금, 자금, 펀드 명칭
	CV_LANGUAGE	- 언어 명칭
	CV_BUILDING_TYPE	- 건축양식 명칭
	CV_FOOD	- 음식 명칭, 음식재료
	CV_DRINK	- 음료수, 술 명칭
	CV_CLOTHING	- 의복/섬유 명칭
	CV_POSITION	- 직위/직책 명칭, 스포츠 포지션
	CV_RELATION	- 인간 관계 명칭
	CV_OCCUPATION	- 직업 명칭
	CV_CURRENCY	- 통화 명칭
	CV_PRIZE	- 상과 훈장
	CV_LAW	- 법/법률 명칭
	CV_FOOD_STYLE	- 음식 종류
ANIMAL (AM)	AM_OTHERS	- 기타 동물 명칭
	AM_INSECT	- 곤충
	AM_BIRD	- 조류
	AM_FISH	- 어류
	AM_MAMMALIA	- 포유류
	AM_AMPHIBIA	- 양서류
	AM_REPTILIA	- 파충류
	AM_TYPE	- 동물 분류 명칭
	AM_PART	- 동물 몸의 한 부분(신체부위) 명칭
PLANT (PT)	PT_OTHERS	- 기타 식물 명칭
	PT_FRUIT	- 과일 이름 (식용 가능한 식물의 열매)
	PT_FLOWER	- 꽃 이름
	PT_TREE	- 나무 이름
	PT_GRASS	- 풀 이름
	PT_TYPE	- 식물 유형 명칭
	PT_PART	- 식물의 한 부분에 대한 명칭
QUANTITY (QT)	QT_OTHERS	- 기타 수량표현
	QT_AGE	- 나이
	QT_SIZE	- 크기/넓이
	QT_LENGTH	- 길이/거리/높이
	QT_COUNT	- 개수/빈도
	QT_MAN_COUNT	- 인원수
	QT_WEIGHT	- 무게
	QT_PERCENTAGE	- 백분율, 비율, 농도

	QT_SPEED	- 속도
	QT_TEMPERATURE	- 온도
	QT_VOLUME	- 부피
	QT_ORDER	- 순서, 순차적 표현
	QT_PRICE	- 금액
	QT_PHONE	- 전화번호
	QT_SPORTS	- 스포츠 관련 수량표현
	QT_CHANNEL	- TV/라디오 채널번호
	QT_ALBUM	- 앨범 관련 수량 표현
	QT_ZIPCODE	- 우편번호
STUDY_FIELD (FD)	FD_OTHERS	- 학문분야 및 학파
	FD_SCIENCE	- 과학 학문분야
	FD_SOCIAL_SCIENCE	- 사회과학 학문분야 및 학파 - 정치/경제/사회와 관련된 분야
	FD_MEDICINE	- 의학관련 학문분야 및 분과
	FD_ART	- 예술관련 학문분야 및 학파
	FD_PHILOSOPHY	- 철학관련 학문분야 및 학파
THEORY (TR)	TR_OTHERS	- 기타 이론/법칙/원리
	TR_SCIENCE	- 과학 관련 이론/법칙/방식/양식/체계/학설
	TR_SOCIAL_SCIENCE	- 사회과학 이론/법칙/방법/원리/사상, 정치사상
	TR_ART	- 예술관련 이론/법칙/방식/양식, 예술사조
	TR_PHILOSOPHY	- 철학 이론/사상
	TR_MEDICINE	- 의학 요법/처방, 의학 진단법
EVENT (EV)	EV_OTHERS	- 기타 사건/사고 명칭, ~사태
	EV_ACTIVITY	- 사회운동 및 선언
	EV_WAR_REVOLUTION	- 전쟁/혁명
	EV_SPORTS	- 스포츠/레저 관련 행사
	EV_FESTIVAL	- 축제 명칭
MATERIAL (MT)	MT_ELEMENT	- 원소명
	MT_METAL	- 금속물
	MT_ROCK	- 암석 명칭
	MT_CHEMICAL	- 모든 화학물질을 나타냄
TERM (TM)	TM_COLOR	- 색
	TM_DIRECTION	- 방향
	TM_CLIMATE	- 기후지역 명칭
	TM_SHAPE	- 모양/형태 명칭
	TM_CELL_TISSUE	- 세포/조직/기관 명칭 (외부에서 안 보이는 조직/기관)
	TMM_DISEASE	- 증상/증세/질병
	TMM_DRUG	- 약/약품명

	TMI_HW	- IT 하드웨어 용어
	TMI_SW	- IT 소프트웨어 용어
	TMI_SITE	- URL 주소
	TMI_EMAIL	- 이메일주소
	TMI_MODEL	- 각종 제품의 세부 모델명, 부품류
	TMI_SERVICE	- IT 서비스 용어
	TMI_PROJECT	- 프로젝트 명칭
	TMIG_GENRE	- 게임 장르
	TM_SPORTS	- 스포츠/레저 용어 (기술/규칙 명칭)

- 새로운 의미 있는 용도나 언어현상의 출현 또는 발견으로 새 범주의 태그 설정이 필요할 경우에는 이 체계를 유지하면서 계층적으로 새로운 하위분류를 추가하여 점증적인 방식으로 다듬어 나갈 수 있으므로, 이 분류체계는 개방적이고 수용적인 기본 틀을 제공하는 것을 목적으로 하고 있다.
- 본 권고안은 개체명 태그세트와 개체명 태깅 지침의 표준안 마련에 목적이 있으며, 기본적인 태깅 지침은 세분류를 기준으로 작성하였다. 하지만, 본 권고안의 사용자가 필요에 따라서 대분류와 세분류를 선택해서 사용할 수 있도록 대분류에도 개체명 태그를 설정하였다.
- 세부분류 개체명에서 몇 가지 고려할 사항으로는 우선, 일반적으로 개체명은 고유명사만을 대상으로 한다는 것이다. 여기서 고유명사란 실 세계에 유일하게 존재하는 특정 인스턴스의 명칭을 나타낸다. 사전적으로 고유명사는 특정한 사물 및 사람의 고유 기호나 명칭을 의미한다. 하지만, 관계(CV_RELATION)와 같은 특정 태그의 경우에는 ‘아버지’, ‘어머니’, ‘자식’과 같은 일반명사도 세부분류 개체명에 포함되어 있다.
- 두 번째로 고려할 사항은 세부분류 개체명 태깅 범위에 관련된 문제점으로 특정 개체명은 태깅 대상의 범위가 모호한 특징을 갖는다. 일례로 “대전 유성구 신성동”의 경우 <LC_OTHERS:대전 유성구 신성동>의 태깅과 <LCP_CITY:대전> <LCP_COUNTY:유성구> <LCP_COUNTY:신성동>의 태깅이 가능하다. 여기서 본 지침서는 두 번째 유형을 선택한다. 다른 예로 “미국 컬럼비아 대학”의 경우 <LCP_COUNTRY:미국> <OGG_EDUCATION:컬럼비아 대학>으로 태깅해야 한다. 그러나 “한국 해양 대학”처럼 국가명이 고유명칭에 포함되어 있는 경우에는 이를 전부 포함하여 <OGG_EDUCATION:한국 해양 대학>으로 태깅해야 한다.
- 세 번째는 태깅 모호성 문제이다. ‘경복궁’의 경우 문화재(AF_CULTURAL_ASSE)와 빌딩(AF_BUILDING)에 모두 해당된다. 본 권고안은 중복태깅을 허용하지 않는다. 따라서, 문맥을 고려하여 가장 알맞은 1가지로 태깅을 해야 한다. 이와 상기한 문제점들에 대하여는 태깅 말뭉치 구축 지침에서 태그세트 별로 세부적으로 지침을

보인다.

6. 태깅 말뭉치 구축 지침

본 문서는 각 개체명 태그세트에 대한 태깅 말뭉치 구축에 대한 지침도 제시한다. 세 부분류 개체명의 명확한 정의, 태깅 정책, 태깅 시 고려될 사항, 예상되는 문제점 등에 대해 지금부터 자세히 기술한다.

6.1. 인물(PERSON)

PERSON은 실존인물, 신화속 인물, 게임/소설의 가상 캐릭터 등을 나타내며, 축약어로 PS(PerSon)으로 표시한다.

6.1.1 PS_NAME

- 정 의
 - 인명 / 인명 별칭
 - 실존 인물
 - 신화/종교 등장인물
 - 게임/소설의 가상 캐릭터 인물
- 예 제
 - 실존인물
 - 흥선대원군, 해원 신윤복, 현종, 프랭클린 루스벨트, 콘스탄티누스, 콘스탄티누스 1세, W.가이거, S.A.키르케고르, F.M.도스토예프스키, 마르쿠스 아우렐리우스
 - 가수 인물: 소녀시대, 아이유
 - 배우 인물: 원빈, 한가인, 장동건
 - 감독 인물: 봉준호, 박찬욱
 - 신화/종교 등장인물 예제
 - 티에스테스, 보현보살
 - 아스페르스, 신농, 엔키두, 웅녀, 전륜성왕, 아가멤논
 - 게임/소설의 가상 캐릭터 인물 예제
 - 짱구, 심슨, 슈퍼맨, 배트맨
- 주 의
 - 해당 사항 없음

6.2. 학문분야(STUDY_FIELD)

STUDY_FIELD는 학문 분야를 나타내며, 축약어로 FD(study Field)로 표시한다.

세부 범주로는 FD_SCIENCE, FD_SOCIAL_SCIENCE, FD_MEDICINE, FD_ART, FD_PHILOSOPHY 다섯가지로 구성되어 있다. 이외의 학문 분야는 FD_OTHERS 로 명시한다. 더불어 STUDY_FIELD는 학파를 포함한다. 과학, 사회과학, 의학, 예술, 철학 관련 학파 및 유파들은 STUDY_FIELD의 세부 범주로 명시된다.

학문 분야는 위키백과의 분류체계로 다음과 같은 분류체계를 기준으로 분류된다.

위키 대분류	위키 세분류	개체명 세부분류
인문학	언어학	FD_SOCIAL_SCIENCE
	역사학	
	종교학	
	철학	FD_PHILOSOPHY
	문예학	FD_ART
	음악사학	
	미술사학	
	공연예술학	
사회과학	정치학	FD_SOCIAL_SCIENCE
	경제학	
	사회학	
	행정학	
	고고학	
	인류학	
	지리학	
	심리학	
	문헌정보학	
자연과학	물리학	FD_SCIENCE
	생명과학	
	우주과학	
	지구과학	
	화학	
응용과학	경영학	FD_SOCIAL_SCIENCE
	교육학	
	신학	
	언론학	
	법학	
	의학	FD_MEDICINE
	공학	FD_SCIENCE
	농학	FD_OTHERS
	군사학	FD_OTHERS
	가정학	FD_OTHERS

	건강과학	FD_OTHERS
	교통학	FD_OTHERS
	도서관학	FD_OTHERS
	박물관학	FD_OTHERS
	삼림학	FD_OTHERS
	스포츠과학	FD_OTHERS

6.2.1 FD_SCIENCE

- 정 의
 - 과학관련 학문 분야 및 과학 학파
 - 과학분야는 크게 순수과학과 기술과학으로 분류된다. 순수과학이란 엄밀한 뜻의 자연과학을 응용과학에 대응해서 쓰는 말이다. 순수과학의 종류로는 물리학, 생명과학, 수학, 지구과학, 화학등이 있다. 이에 반해 기술과학이란 일종의 응용과학을 말하는 것으로 건축공학, 금속재료공학, 기계공학, 기타 공학, 무기 및 병기공학, 산업공학, 수의학, 원자력공학, 자동차공학, 전기공학, 전자공학, 조선해양공학, 토목건설공학, 항공우주공학, 화학공학, 환경공학 등이 이에 해당한다.
- 예 제
 - 순수과학
 - 물리학, 생명과학, 수학, 지구과학, 화학, 생물물리화학, 소립자물리학, 뉴턴 역학, 양자역학, 생물학, 유전학, 응용지질학, 천문학 등
 - 기술과학
 - 건축공학, 금속재료공학, 기계공학, 기타 공학, 무기 및 병기공학, 산업공학, 수의학, 원자력공학, 자동차공학, 전기공학, 전자공학, 조선해양공학, 토목건설공학, 항공우주공학, 화학공학, 환경공학 등
- 주 의
 - 해당 사항 없음

6.2.2 FD_SOCIAL_SCIENCE

- 정 의
 - 사회과학 관련 학문 분야 및 학파
 - 사회과학이란 인간 사회의 여러 현상을 과학적, 체계적으로 연구하는 모든 경험과학을 말한다. 사회과학의 분야로는 경영, 경제, 교육, 국방 및 군사, 뉴스와 미디어, 법, 사회, 정치, 통계, 행정 등이 포함된다.
- 예 제
 - 경영학, 국제경영학, 신고전학파
 - 사회인류학, 기능학파, 뒤르켐학파, 통계학
 - 인류학, 사회학, 경제학, 정치학, 심리학, 인문지리학, 법학, 응용사회과학, 경영학, 고고학, 행정학, 문헌정보학, 역사학, 언어학, 종교학, 교육학, 신학, 언론학,

가정학

- 주 의
 - 해당 사항 없음

6.2.3 FD_MEDICINE

- 정 의
 - 의학관련 학문 분야 및 학파
 - 의학이란 인간을 질병으로부터 구하고 건강법을 모색하는 학문이다.
 - 의학 학문분야는 양학과 한의학을 포함하여 각 학문별 세부 학문 명칭을 나타낸다.
- 예 제
 - 학문분야
 - 의학, 병리학, 보건의료학, 신경과학, 가정의학, 감각생리학, 기초의학, 생리학, 해부학, 생리체학, 기생충학, 면역학, 미생물학, 병리학, 생화학, 약리학, 의료정보학, 의사학, 법의학, 치의학, 약학, 한의학, 기상의학, 구강위생학
 - 보건학, 약학, 내분비학, 병리학, 열대의학
 - ~의학과
 - 분과
 - 비뇨기과, 신경과, 가정의학과, 간호학과, 내과, 소아청소년과, 정신과, 피부과, 재활의학과, 안과, 이비인후과, 외과, 산부인과, 영상의학과, 핵의학과, 방사선종양학과
 - ~외과
- 주 의
 - 비뇨기 등은 FD_MEDICINE이 아님
 - 의료 학과(분과)의 경우에는 FD_MEDICINE으로 태깅한다.

6.2.4 FD_ART

- 정 의
 - 예술관련 학문분야 및 학파(유파)
 - 예술관련 학문분야는 특정 학문 명칭이라기 보다는 예술의 분류 명칭이라 볼 수 있다. 무용, 미술, 연극, 영화, 문학, 사진, 음악, 현대예술의 각 세부 예술 형태의 분류 명칭이 FD_ART에 해당한다. 미술의 경우 공예, 디자인, 서예, 조각, 판화, 회화 등이 이에 해당한다.
- 예 제
 - 학문분야
 - 문예학, 음악사학, 미술사학, 공연예술학
 - 학파
 - 인상파, 점묘파, 우키요에 화파, 신경향파

- 예술 분류
 - 현대음악, 동양화, 서양화, 한국화, 조각, 공포영화, 과학영화, 수채화, 현악4중주
 - 소설, 뮤지컬, 음악, 영화, 연극, 미술, 팝송, 시
- 주 의
 - 해당 사항 없음

6.2.5 FD_PHILOSOPHY

- 정 의
 - 철학관련 학문분야 및 학파(유파)
 - 철학이란 인생, 세계 등등에 대해 연구하는 학문으로 크게 논리학, 동양철학, 서양철학, 심리학, 윤리학 등으로 분류된다.
- 예 제
 - 학문분야
 - 교육철학, 형이상학, 윤리학, 정치철학, 과학 철학, 언어철학, 논리학, 예술철학, 미학, 윤리학, 사회철학, 스콜라철학, 주기철학, 과학철학, 성리학, 심리학
 - 학파
 - 스콜라학파, 플라톤 학파, 신플로이토파
- 주 의
 - ‘빈 학단’과 같은 철학 단체는 FD_PHILOSOPHY 대상이 아님
 - "∼ 주의"의 경우에는 이론(THEORY) 분류에 해당되고, "∼주의 학파"의 경우에 학문분야(FIELD) 분류에 해당됨

6.2.6 FD_OTHERS

- 정 의
 - 기타 학문분야 및 학파
 - FD_SCIENCE, FD_SOCIAL_SCIENCE, FD_MEDICINE, FD_ART, FD_PHILOSOPHY의 세부 학문분야 이외의 학문분야 및 학파
- 예 제
 - 농학, 군사학, 가정학, 건강과학, 교통학, 도서관학, 박물관학, 산림학, 스포츠과학
- 주 의
 - 해당 사항 없음

6.3. 이론(THEORY)

THEORY는 특정 이론, 법칙, 원리 등의 명칭을 의미하며, 축약어로 TR(TheoRy)로 표시한다. THEORY는 명확히 기술될 수 있는 고유의 이론, 법칙, 원리만을 말하며, 일반적인

용어들(예: 사회법칙, 법이론 등)은 해당하지 않는다. 세부 분류로는 TR_SCIENCE, TR_SOCIAL_SCIENCE, TR_ART, TR_PHILOSOPHY, TR_MEDICINE으로 분류 된다. 이 세부 분류를 벗어나는 이론, 법칙 명칭은 TR_OTHERS에 포함된다.

이론과 관련된 몇 가지 공통적인 주의사항은 다음과 같다.

- THEORY는 ‘~적’의 형태로 기술될 수 있다. 이 경우 ‘~’는 THEORY로 간주 한다.
- THEORY는 ‘~ 사상, ~ 원리, ~ 이론’의 형태로 기술될 수 있다. 이 경우 ‘~ 사상, ~ 원리, ~ 이론’을 포함해서 THEORY로 간주한다.
- THEORY는 ‘천황송배’ 사상의 형태로 기호와 함께 쓰일 수 있다. 이 경우 기호를 포함해서 THEORY로 간주한다.

6.3.1 TR_SCIENCE

- 정 의
 - 순수과학: 과학 관련 이론/법칙/방식/양식/체계/학설
 - 기술과학: 기계/장치/시스템/기술 관련 이론/법칙/방식
 - 작용/현상은 TR_SCIENCE의 대상으로 보지 않음
- 예 제
 - 순수과학
 - 특수상대성이론, 우열의 법칙, 양자론, 멘델법칙, 드무아브르의 정리, 갈릴레이-뉴턴의 상대성원리
 - 21세기형수학공식, 이진법, 2진수, 3원질설
 - 기술과학
 - 원심분리법, 기체확산법, 원심분리법, 열확산법, 컨베이어 라인(conveyer line) 시스템
 - 진공건조, 이온주입법, 공랭식, 수랭식
- 주 의
 - 고유의 명칭을 갖고 있는 특정 과학 이론만이 태깅 대상이다.
 - 다음은 부적절한 TR_SCIENCE의 예제들이다.
 - 서로 등속도로 운동하는 관측자에 대하여 전자기파의 이론을 포함한 모든 물리법칙이 같은 형식으로 기술되도록 정식화되어 있다.
 - 현재는 양자역학과 더불어 현대물리학, 특히 소립자물리학이나 우주론의 지도적 원리로 간주되고 있다.
 - X선의 성질과 작용: X선은 사진작용·형광작용·이온화작용, 진공 상태에서는 빛과 같은 속도로 진행한다는 성질이 있고, 그 외에도 입자와 같이 회절한다.
 - 중력의 크기는 물체의 질량에 비례하며, 질량 1 g의 물체에는 대체로 980 dyn의 중력이, xg의 물체에는 980 dyn의 x배의 중력이 미친다. 보통 물체의 무게라고 하는 것은 이 힘을 가리킨다.
 - 빙식작용은 크레바스(베르크슈르트라고도 한다)의 밑바닥에서 융빙수(融氷水)

의 동결에 의해 행하여지는 파쇄작용과, 빙하의 유동에 따라 이루어지는 연마작용의 두 가지 방법으로 기반암석(基盤岩石)을 침식한다.

- 그러므로 관성력은 가속도 운동을 하고 있는 관측자에게만 느껴진다. 관성력은 힘의 원인인 물체가 없는 가상의 힘이므로 작용·반작용의 법칙이 성립하지 않는다.

6.3.2 TR_SOCIAL_SCIENCE

- 정 의
 - 사회과학 이론/법칙/방법/원리/사상, 정치사상
- 예 제
 - 민주주의, 공산주의, 다원적 민주주의, 법치주의, 평등주의, 관료주의, 민족주의, 자본주의
 - 무작위추출법, 노동가치설, 대화법
- 주 의
 - 개체명 태깅 기본 원칙에 따라 최장 유형을 태깅 대상으로 한다. 아래의 예에서 ‘신고전학파 성장론’이 이론 명칭이 된다. 여기서 ‘신고전학파’라는 FD_SOCIAL_SCIENCE가 포함되어 있으나 최장 개체명 태깅 원칙에 따라 ‘신고전학파 성장론’을 TR_SOCIAL_SCIENCE로 태깅한다.
 - 본문 예제: 신고전학파 성장론의 경우에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 신고전파종합 (neoclassical synthesis)의 경우에는 경제적 복지의 달성수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.

6.3.3 TR_ART

- 정 의
 - 예술관련 이론/법칙/방식/양식, 예술사조
- 예 제
 - 에로티시즘, 원근법, 대기원근법, 선원근법, 구름법, 물골법
 - 대위법, 데카당스, 문자기보법, 환유법, 제유법
 - 고딕양식, 고전주의, 유량악보
- 주 의
 - 해당 사항 없음

6.3.4 TR_PHILOSOPHY

- 정 의
 - 철학 이론/사상
- 예 제

- 조작주의, 주지주의, 사회주의, 개인주의, 구조주의, 토마스주의, 경험주의, 합리주의, 공리주의, 현실주의
- 니힐리즘, 이데아설, 인식론, 노장사상, 이기론, 존재론, 관념론
- 유가, 도가, 묵가, 법가, 명가, 음양가, 종횡가, 병가
- 주 의
 - “플라톤 학파” 등의 철학 유파는 FD_PHILOSOPHY이다.

6.3.5 TR_MEDICINE

- 정 의
 - 의학 요법/처방, 의학 진단법
- 예 제
 - 팔성혈 자세교정법, 인공호흡, 요혈자극법, 뜸, 바이오피드백
 - 발열요법, 견인요법, 보법, 대기요법, 공기요법
 - 수지력 테스트, X선검사
 - 맥진, 내시경검사, 혈액검사, 검안, 문진
- 주 의
 - 해당 사항 없음

6.3.6 TR_OTHERS

- 정 의
 - 기타 이론/법칙/원리
 - TR_SCIENCE, TR_SOCIAL_SCIENCE, TR_ART, TR_PHILOSOPHY, TR_MEDICINE의 세부 이론 분야 이외의 이론/법칙/원리
- 예 제
 - 샤머니즘, 원칙주의
- 주 의
 - 해당 사항 없음

6.4. 인공물(ARTIFACTS)

ARTIFACTS는 인공물을 나타내며, 축약어는 AF(ArtiFacts)로 나타낸다. 인공물이란 사람에 의해 창조된 대상을 말하며 문화재, 건물, 악기, 도로, 무기, 운송수단, 작품명 등이 이에 해당한다. 유의할 점은 ARTIFACTS 내부에 포함된 세부범주에 해당되지 않는 인공물은 태깅 대상에서 배제한다.

인공물의 공통적인 주의 사항은 다음과 같다.

- ARTIFACTS 세부 태그 중 일부는 서로 중복 태깅이 가능할 경우가 있다. 개체명태깅은 단일 개체명태깅 원칙을 갖는다. 따라서, 중복태깅 가능한 개체명 어휘가

출현했을 경우 우선순위를 정하여 유일태그를 할당해야 한다. 다음은 이에 대한 정리이다.

중복유형		예제
개체명태그1	개체명태그2	
AF_CULTURAL_ASSET(O)	AF_BUILDING	경복궁, 보현사 등
AF_CULTURAL_ASSET	AF_WORKS(O)	고려사, 신증동국여지승람, 동의보감, 인왕제색도
AF_CULTURAL_ASSET	LC_TOUR(O)	천연기념물(예: “단양영천리측백나무숲”, “담양관방제림”)

※ O로 표시된 태그의 우선순위가 높음

6.4.1 AF_CULTURAL_ASSET

- 정 의
 - 문화재 명칭
- 예 제
 - 경복궁, 보현사, 경주 서악리 마애석불상, 불국사, 석굴암, 관상감 측우대, 문양전
- 주 의
 - 문화재는 다양한 형태를 지닌다. 역사적인 건물, 서적, 작품 모두 문화재라 볼 수 있다. 그러나 ARTIFACTS의 세부 개체명에 세부적 사항이 대부분 기술되어 있다. 따라서, 개체명은 각 세부 분류에 준한다. 그러나 AF_BUILDING의 경우 AF_CULTURAL_ASSET에 해당한다면 문화재로 간주한다. 각 절, 고궁 등은 모두 문화재로 간주한다.

6.4.2 AF_BUILDING

- 정 의
 - 건축물/토목건설물, 운동장이름, 아파트, 다리, 등대, 분수
- 예 제
 - 대한생명63빌딩, 구대한의원 본관, 서해대교, 대청댐, 충주댐, 세계무역센터, 동대문운동장
- 주 의
 - 도로 명칭은 AF_ROAD에 해당된다. 빌딩, 다리, 댐 명칭 등이 AF_BUILDING에 해당한다.
 - 문화재인 절은 AF_CULTURAL_ASSET에 포함된다.

6.4.3 AF_MUSICAL_INSTRUMENT

- 정 의
 - 악기 명칭
- 예 제
 - 기타, 캐스터네츠, 편경, 팽과리, 대금, 해금, 파이프오르간, 오보에
- 주 의
 - 해당 사항 없음

6.4.4 AF_ROAD

- 정 의
 - 도로/철로 명칭
- 예 제
 - 시베리아철도, 경부고속도로, 강원선, 호남선, 경원선, 경의선, 세종로, 올림픽대로, 충북선
- 주 의
 - 해당 사항 없음

6.4.5 AF_WEAPON

- 정 의
 - 무기 명칭
- 예 제
 - 기관단총, 시한폭탄, 곡사포, 고정포, 승차총통, 탄도미사일, 최루가스, 스마트폭탄
 - 무기 유형(일반명사) 포함: 도끼, 칼, 전투기
- 주 의
 - 특정 무기 명칭과 무기 유형 명칭 모두 해당된다.

6.4.6 AF_TRANSPORT

- 정 의
 - 교통수단/자동차/선박 모델 및 유형
 - 운송수단 및 놀이기구
- 예 제
 - 갈라테아호, 그레이트브리튼호, 경비정, 가스터빈자동차, 메르세데스벤츠, 자기부상열차, 나룻배, 경자동차, 탈리스
 - 교통수단 유형(일반명사) 포함: 자동차, 배, 비행기
- 주 의
 - 해당 사항 없음

6.4.7 AF_WORKS

- 정 의
 - 기타 작품명
- 예 제
 - 주천도분환, 성구백각환
- 주의
 - '조선일보'와 같은 신문사는 OGG_MEDIA에 해당됨
 - AFW 계열의 세부 작품명에 포함되지 않는 작품명의 경우 AF_WORKS에 해당한다.

6.4.8 AFW_DOCUMENT

- 정 의
 - 도서/서적 작품명
 - 지리서, 지도
 - 의학관련 서적
 - 종교서적, 각 종교의 경전
 - 철학관련 서적
 - 시/소설/희곡 등의 문학 작품명
 - 역사서적 (예: 중종실록, 고려사)
 - 기타 서적 (예: 뽀뽀한 영철영어)
- 예 제
 - 지리서, 지도
 - 신증동국여지승람, 대동여지도, 각도여지도, 동국여지승람, 아방강역고, 택리지, 혼일강리역대국도지도, 곤여만국전도
 - 의학 서적
 - 동의보감, 의림비서
 - 본초강목, 의방신감, 삼화자향약방, 의약감
 - 종교 서적
 - 한역대장경, 장아함경, 중아함경, 잡아함경, 증일아함경, 흥정역성서, 개역성서, 코란, 요한의 묵시록, 아베스타
 - 철학 서적
 - 존재와 시간, 노자도덕경, 범주론, 명제론, 토정비결, 장자, 예기
 - 시/소설/희곡 등의 문학 작품명
 - 돈카를로스, 쿠오바디스, 구스타프 3세, 수상기, 반지의 제왕, 난장이가 쏘아올린 작은 공, 모란이 피기까지는
 - 세부 예제: 역사 서적
 - 중종실록, 고려사)
 - 세부 예제: 기타 서적

- 뻔뻔한 영철영어
- 주 의
 - 해당 사항 없음

6.4.9 AFW_PERFORMANCE

- 정 의
 - 춤/무용 작품명 및 춤 종류 (예: 백조의 호수, 재즈댄스)
 - 연극/공연명/가극 (예: 캐츠, 돈카를로스, 피바다)
- 예 제
 - 춤/무용 작품명
 - 농악무, 동래학춤, 바라춤, 간척무, 검무, 사교댄스, 에어로빅댄스, 백조의 호수, 세에라자드, 재즈댄스
 - 연극/공연명
 - 돈카를로스, 피바다, 밀림아 이야기하라, 명성황후, 캐츠, 마이 페어 레이디
- 주 의
 - 특정 춤/무용 창작 작품명 뿐 아니라, 춤 종류도 포함된다.

6.4.10 AFW_VIDEO

- 정 의
 - 영화 작품명
 - TV 프로그램 이름 (예: 아빠 어디가, 9시 뉴스, 1박 2일)
 - 애니메이션 작품명
- 세부 예제
 - 영화 작품명
 - 사랑방 손님과 어머니, 다이하드, 쉬리, 박하사탕, 바람의 검심, 올랜도, 마리 이야기, 취화선
 - TV 프로그램명
 - 아빠 어디가, 9시 뉴스, 1박2일
- 주 의
 - 해당 사항 없음

6.4.11 AFW_ART_CRAFT

- 정 의
 - 미술 작품명(그림/미술품)
- 예 제
 - 생각하는 사람, 도산도, 도산서원도, 풍류절창, 춘투도, 춘화사계첩, 원앙비보, 키스, 게르니카, 지옥의 문

- 주 의
 - 조각품 등의 조형 작품도 해당됨. 불교에서 나오는 그림명, 불상명도 포함됨

6.4.12 AFW_MUSIC

- 정 의
 - 음악 작품명
- 예 제
 - 운명의 힘, 돈 카를로스, 아이다, 카르멘, 한국환상곡, 보태평, 정대업
- 주 의
 - 오페라 작품도 해당됨

6.4.13 AF_WARES

- 정 의
 - 상품/제품 이름
 - 운동화, 화장품, 여행상품 등
- 예 제
 - 피테라, 하이-믹스, 크리스탈로맨스
- 주 의
 - 다른 개체명 태그셋과 겹칠 경우, 다른 개체명 태그셋을 우선 함
 - TMI_HW, CV_FOOD, CV_DRINK, TMI_MODEL 등으로 태깅 가능하면 그쪽으로 태깅
 - 화장품이나 향수 관련 제품 등은 특별한 태그셋이 없으므로 이 태그를 사용

6.5. 기관(ORGANIZATION)

ORGANIZATION은 기관/단체 명칭을 나타내며, 축약어로 OG(OrGanization)으로 표시한다. 경제기관, 교육기관, 군사기관, 미디어/방송 기관, 스포츠기관, 예술기관, 의료기관, 종교기관, 과학기관, 도서관, 법률기관, 정치/정부기관, 음식 관련 기관, 숙박기관 등으로 세분화된다. 세부 기관에 포함되지 않는 기관/단체들은 OG_OTHERS로 태깅된다.

기관의 공통적인 주의사항은 다음과 같다.

- 학파는 STUDY_FIELD로 분류된다.
- “서울대학 응용미술학과 졸업”에서 서울대학 응용미술학과는 기관/단체명으로 볼 수 없다. “서울대학”만을 기관/단체명으로 한다.
- 특정 기관/단체명과 그 기관/단체의 세부 조직 명칭이 동시에 쓰였을 경우 이를 묶어 전체를 기관/단체명으로 보지 않고 대표 기관/단체명만을 태깅한다.

6.5.1 OGG_ECONOMY

- 정 의
 - 경제 관련 기관/단체
 - 회사/기업
 - 은행, 카드사, 보험사, 증권사
 - 사기업(삼성그룹, 삼성전자, LG그룹, 현대자동차 등)
- 예 제
 - 경제 기관/단체
 - 공정거래위원회, 금융감독원, OECD, 국제통화기금, 석유수출국기구
 - 회사/기업
 - BYC(주), 포드자동차, 불교방송, 한국도로공사, 네슬레, 휴렛패커드, 동양척식주식회사, 한국은행, 비씨카드, LG화재, 동양증권
- 주 의
 - “공정거래위원회 사무처”를 기관명으로 볼 수 없다.
 - 기관/단체명은 공식적인 명칭에 한한다. 하위 조직을 포함해서 기관/단체명으로 하지 않는다.
 - 식물원, 수목원, 수족관, 숲 등은 LC_TOUR로 태깅함
 - 구로공단은 LC_OTHERS에 해당됨
 - 경제 관련 공공기관/정부기관은 OGG_POLITICS에 해당됨 (예: 근로복지공단, 한국산업단지공단)

6.5.2 OGG_EDUCATION

- 정 의
 - 교육 기관/단체
 - 학교, 학원, 어린이집, 훈련소, 학회
- 예 제
 - 홍익대학교, 국제연합교육과학문화기구
 - 국자감, 대한지리학회, 국군간호사관학교, 한인중앙학원
 - 고용훈련소, 논산훈련소
- 주 의
 - ‘유네스코’와 같이 교육, 문화, 과학의 여러 분야와 관련된 기관은 OG_OTHERS로 태깅함
 - 학교 학과의 경우에는 OGG_EDUCATION으로 태깅하지 않음. 개체명 대상이 아님
 - 예) 1954년 <부산대학교:OGG_EDUCATION> 법학과를 졸업하였으며….
 - ~ 훈련소는 모두 OGG_EDUCATION으로 태깅한다.
 - 학회의 경우에는 학회의 종류에 OGG_EDUCATION 이외의 세부적인 분류가 가능하지만, 모든 학회는 OGG_EDUCATION으로 태깅한다.
 - 예) 한국정보과학회: OGG_SCIENCE or OGG_EDUCATION →

OGG_EDUCATION

6.5.3 OGG_MILITARY

- 정 의
 - 군사 기관/단체 및 유형, 테러조직/무장단체
 - 군대 유형(일반명사)도 포함됨: 해군, 공군, 육군, 인민군
- 예 제
 - 아르헨티나군, 대한민국재향군인회, 국방품질관리소, 한미연합사령부, 국방홍보원, 국제연합군
- 주 의
 - '탈레반', '이슬람국가' 등의 무장단체도 OGG_MILITARY에 해당됨

6.5.4 OGG_MEDIA

- 정 의
 - 미디어 기관/단체
 - 방송관련 회사/기업
- 예 제
 - 국제잡지협회, 국제기자연맹, KBS, MBC, TJB, 방송심의위원회, 공익광고협의회, 스카이라이프㈜
- 주 의
 - '조선일보' 등과 같이 신문사(OGG_MEIDA)와 출판물(AF_WORKS)이 혼용되는 경우에는 OGG_MEDIA로 할당
 - '스카이라이프'와 같이 방송관련 기업도 OGG_MEDIA로 할당
 - '불교방송'과 같은 경우에도 종교 관련 방송기관도 OGG_MEDIA에 해당됨

6.5.5 OGG_SPORTS

- 정 의
 - 스포츠 기관/단체
- 예 제
 - 국민체육진흥심의위원회, 국기원, 국민생활체육협의회, 국군체육부대, 한국빙상지도자협의회
- 주 의
 - “~ 대표팀”의 일반 명사도 OGG_SPORTS로 태깅한다
 - 예) 핸드볼대표팀, 대한민국 축구대표팀
 - 1부(리그) 등은 '잉글랜드 3부 리그'로 구체화된 경우에만 OGG_SPORTS로 할당
 - 우승팀, 준우승팀의 일반명사는 태깅 대상에서 제외
 - “~ 리그”의 경우는 EV_SPORTS로 할당

- 예: K리그, 메이저리그, 프리미어리그

6.5.6 OGG_ART

- 정 의
 - 예술 기관/단체
 - 공연장, 전시장, 박물관, 미술관, 극장
- 예 제
 - 대한극장, 구겐하임미술관, 국립중앙박물관, 예술의 전당, 한국민속박물관, 로스 앤젤레스필하모니관현악단, 동양극장
- 주 의
 - 해당 사항 없음

6.5.7 OGG_MEDICINE

- 정 의
 - 의학/의료 기관/단체
 - 병원, 약국 등
- 예 제
 - 세계보건기구, 국립재활원, 경희의료원, 대한간호협회, 의료보험연합회
- 주 의
 - “세계보건기구헌장”과 같이 기관명을 포함하고 있는 복합명사의 경우 기관명을 배제한 어휘가 독립된 어휘로 쓰일 수 있을 경우 태깅 대상으로 한다. 즉, “세계 보건기구/헌장”과 같이 복합명사분해가 가능할 경우 “세계보건기구”는 기관 단체 명으로 본다.

6.5.8 OGG_RELIGION

- 정 의
 - 종교 기관/단체 및 종파
- 예 제
 - 증산교, 구가톨릭교
 - 대한성공회, 한국기독교장로회, 대한불교화엄종, 조선불교청년회
- 주 의
 - “구가톨릭교도”와 같이 종교 기관이 복합명사의 일부이고, 종교 기관만을 독립적 인 의미로 분리하기 힘든 경우에는 OGG_RELIGION으로 태깅하지 않는다.
 - 종교/종파의 교단과 같은 경우에는 OGG_RELIGION으로 태깅하지 않는다.

6.5.9 OGG_SCIENCE

- 정 의
 - 과학 기관/단체
 - 연구소, 연구원
- 예 제
 - 국제전기통신연합, 국립국어연구원, 농지개량조합, 한국전기안전공사, 정보통신연구진흥원
- 주 의
 - “한국정보과학회”와 같은 과학 관련 학회는 OGG_SCIENCE로도 볼 수 있지만, 모든 학회는 OGG_EDUCATION으로 태깅한다.

6.5.10 OGG_LIBRARY

- 정 의
 - 도서관 및 도서관 관련 기관/단체
- 예 제
 - 국립중앙도서관, 바티칸도서관, 알렉산드리아도서관, 국회도서관, 한국도서관협회, 대동서관, 바티칸도서관
- 주 의
 - 해당 사항 없음

6.5.11 OGG_LAW

- 정 의
 - 법률 기관/단체
- 예 제
 - 대한법률구조공단, 광주지방법원, 대한법무사협회, 저작권심의조정위원회, 대한변호사협회, 헌법재판소
- 주 의
 - 법원은 국가기관(OGG_POLITICS)과 법률기관(OGG_LAW)에 모두 해당되는데, OGG_LAW로 태깅한다.
 - 대한제국 법무 전라북도재판소 → <대한제국:LCP_COUNTRY> <법부:OGG_LAW> <전라북도재판소:OGG_LAW>
 - 미 군정청 광주지방법원 → <미:LCP_COUNTRY> <군정청:OGG_POLITICS> <광주지방법원: OGG_LAW >

6.5.12 OGG_POLITICS

- 정 의
 - 정부/행정 기관, 공공기관, 정치기관
 - 대사관, 우체국, 경찰서, 소방서, 동사무소, 전화국, 시청, 군청, 대검, 검찰, 검찰

청

- 공단, 공사, 진흥원
- 예 제
 - 정부/행정 기관
 - 미래부, 정보통신부, 강남경찰서, 성동소방서, 광명6동사무소, 개봉전화국, 광명시청, 영동군청, 국회, 경제기획원, 북방경제정책 실무위원회, 신결조성도감, 기획예산처, 행정자치부, 경찰청, 스웨덴대사관
 - 공공기관
 - 근로복지공단, 사립학교교직원연금공단, 인천국제공항공사, 한국사회적기업진흥원
 - 정치기관
 - 새누리당, 새천년민주연합, 민주당
- 주 의
 - 법원은 국가기관(OGG_POLITICS)과 법률기관(OGG_LAW)에 모두 해당되는데, OGG_LAW로 태깅한다.

6.5.13 OGG_FOOD

- 정 의
 - 음식 관련 업체 및 회사
- 예 제
 - 맥도날드, 김밥천국, 맛나분식
- 주 의
 - 해당 사항 없음

6.5.14 OGG_HOTEL

- 정 의
 - 숙박 관련 업체 및 회사
 - 호텔, 모텔, 민박, 펜션, 콘도, 리조트
- 예 제
 - 신라호텔, 한화콘도, 대명리조트
- 주 의
 - 휴양시설(ex. 마리나 리조트, 무주 리조트, 용평 리조트, 휘닉스파크, 하이원리조트)은 OGG_HOTEL과 LC_TOUR가 모두 가능한데, OGG_HOTEL로 태깅한다.
 - 호텔 내 존재하는 룸 이름(ex. 주니어 스위트룸)은 개체명 대상(OGG_HOTEL)이 아니다.

6.5.15 OG_OTHERS

- 정 의
 - 기관/단체의 세부 분류에 해당되지 않는 나머지 모든 기관/단체
- 예 제
 - 유색인종향상전국협회, 국제소비자연맹
 - 육영재단, 녹색연합, 참여연대, 한국정신대문제대책협의회
 - 게임 관련 기관: 길드, 클랜, 파티 (게임회사는 OGG_ECONOMY에 해당됨)
- 주 의
 - 해당 사항 없음

6.6. 지역(LOCATION)

LOCATION은 지역/장소와 지형/지리명칭 등을 모두 포함한다. 지역/장소는 LCP(LoCation Place) 계열로 표시하고, 지형/지리 명칭은 LCG((LoCation Geography) 계열로 표시한다. 기타 지역으로 관광명소는 LC_TOUR, 천체 명칭은 LC_SPACE로 태깅한다.

지역의 공통적인 주의사항은 다음과 같다.

- 지역/장소
 - 지역명으로 주로 공식적 지역명칭과 행정구역 명칭을 나타낸다.
 - 지역명은 세부적으로 분리하여 태깅한다.
 - ✓ 충청남도 예산군 덕산면 상가리 → <충청남도:LCP_PROVINCE> <예산군:LCP_COUNTY> <덕산면:LCP_COUNTY> <상가리:LCP_COUNTY>
 - 중복태깅이 가능할 경우 다음 규칙을 따른다.

중복유형	예제	태깅방식
LC_TOUR & LC_PLACE	경포	관광지일 경우 LC_TOUR에 우선순위

- 지형/지리 명칭
 - 국가명이나 행정구역명과는 별개인 산, 강과 같은 자연에 대한 지리명칭
 - 지형의 유형은 포함되지 않고, 구체적인 지형 명칭만 해당된다.

6.6.1 LCP_COUNTRY

- 정 의
 - 국가명
- 예 제
 - 대한민국, 일본, 중국, 미국, 영국, 터키, 그리스, 짐바브웨, 케냐
- 주 의
 - 해당 사항 없음

6.6.2 LCP_PROVINCE

- 정 의
 - 도, 주 지역명
- 예 제
 - 경기도, 강원도, 캘리포니아주, 애리조나주, 충청남도, 퀘벡주, 슈비츠주, 네바다주
- 주 의
 - 해당 사항 없음

6.6.3 LCP_COUNTY

- 정 의
 - 군, 면, 읍, 리, 동 등과 같은 세부 행정구역명
- 예 제
 - 영월군, 도암면, 횡계리, 노암동, 나주군, 완도군, 담양읍, 춘포면, 영남리
- 주 의
 - 해당 사항 없음

6.6.4 LCP_CITY

- 정 의
 - 도시명
- 예 제
 - 강릉, 부산, 태백, 뉴욕, 맨체스터, 덴버, 요코하마, 부에노스아이레스, 이스탄불
- 주 의
 - 국가의 수도명은 LCP_CAPITALCITY으로 태깅한다.

6.6.5 LCP_CAPITALCITY

- 정 의
 - 수도명
- 예 제
 - 서울, 북경, 베이징, 도쿄, 런던, 아테네, 워싱턴, 캔버라, 파리, 브라질리아
- 주 의
 - 해당 사항 없음

6.6.6 LCG_RIVER

- 정 의
 - 강/호수/육수 이름

- 예 제
 - 섬진강, 대동강, 한강, 충주호, 동강, 춘천호, 아산호
- 주 의
 - 해당 사항 없음

6.6.7 LCG_OCEAN

- 정 의
 - 해양/바다 명칭
- 예 제
 - 남극해, 북극해, 데이비스해, 아문센해, 동해, 태평양, 인도양, 산호해
- 주 의
 - 해당 사항 없음

6.6.8 LCG_BAY

- 정 의
 - 반도/만의 명칭
- 예 제
 - 도쿄만, 라몬만, 라오둥만, 라오둥반도, 아산만, 울산만, 변산반도, 발칸반도, 이베리아반도
- 주 의
 - 해당 사항 없음

6.6.9 LCG_MOUNTAIN

- 정 의
 - 산맥/산의 명칭
- 예 제
 - 라우산, 라웅산, 라키화산, 래밍턴산
 - 태백산맥, 차령산맥, 계룡산, 금강산, 대청봉
- 주 의
 - 해당 사항 없음

6.6.10 LCG_ISLAND

- 정 의
 - 섬/제도 명칭
- 예 제
 - 나이팅게일섬, 노시베섬, 로드리게스섬, 로스제도, 마스카렌제도

- 대난지도, 대부도, 백령도
- 주 의
 - '제주도'는 섬의 이름이기도 하지만, 행정구역으로 '제주특별자치도'를 의미하기도 하므로 LCP_PROVINCE로 할당한다.

6.6.11 LCG_CONTINENT

- 정 의
 - 대륙 명칭
- 예 제
 - 아메리카, 아시아, 오세아니아, 유라시아, 유럽, 북아메리카, 남아메리카, 아프리카
- 주 의
 - 해당 사항 없음

6.6.12 LC_TOUR

- 정 의
 - 관광명소
 - 온천, 해변, 관광지+해안, 식물원, 수목원, 정원, 공원, 세트장, 광장, 테마파크, 농원/농장, 폭포, 계곡, 동굴, 야영장, 체험장, 수련장, 샘터, 약수터
- 예 제
 - 간현유원지, 대성산유원지, 디즈니월드, 디즈니랜드, 서울랜드
 - 다도해 해상국립공원, 에버랜드, 예송리해수욕장, 계룡산국립공원
 - 해운대해안, 양남해변, 백암온천, 여미지식물원, 국립수목원, 미소정원, 울산바위, 노안청소년야영장, 이안숲속, 용정약수터
- 주 의
 - ~사막, ~오아시스는 LCP_TOPOGRAPHY나 LC_TOUR가 아니라LC_OTHERS에 해당됨

6.6.13 LC_SPACE

- 정 의
 - 항성, 행성 명칭
 - 위성/유성 및 혜성
 - 별자리, 성운 명칭
- 예 제
 - 고리성운, 견우성, 게성운, 태양, 시리우스, 말머리성운, 장미성운, 미자르, 거문고자리 유성군, 단주기혜성, 모어하우스혜성, 수성, 금성, 토성, 세종, 명왕성, 미란다, 헬리혜성, 거문고자리, 게자리, 고래자리, 고물자리., 사자자리, 나침반자리,

물병자리, 오리온자리

- 주 의
 - 특정 천체에 대한 지칭만을 대상으로 하고, 일반적 천체유형은 배제
 - 예) “헬리헤성”은 LC_SPACE이지만, “헤성”은 개체명(LC_SPACE)이 아님

6.6.14 LC_OTHERS

- 정 의
 - 기타 지역/장소
 - LC계열의 세부유형이 아닌 기타 장소
 - 지하철/기차역, 공항, 버스터미널, IC, 시장, 발전소, 정류장, 단지, 항구, 골짜기
- 예 제
 - 북해유전, 북자파라나 유전
 - 가로구릉, 기아나고지, 나가구릉
 - 미시시피삼각주, 희망봉, 동아프리카대지구대
- 주 의
 - 해안, 해변(ex. 서해안, 해운대해안, 양남해변): 해변은 LC_TOUR, 해운대해안 같이 관광지 해안은 LC_TOUR, 서해안/남해안 등은 LC_OTHERS
 - 제주항/목포항: 만/해안이 아닌 항의 경우는 LC_OTHERS
 - 슬로프명(이스턴 슬로프 루키 슬로프): LC_OTHERS
 - 시장명 (ex. 자갈치시장, 어시장, ...): LC_OTHERS
 - ~검문소(배둔검문소)/무기제작소/군기창고/군사보호구역/군사훈련장: 개체명인 경우에는 LC_OTHERS에 해당됨
 - ‘서양’, ‘동양’, ‘동구’, ‘서구’, ‘북방’, ‘서방’, ‘남방’, ‘동방’과 같은 일반명사도 대상임

6.7. 문명(CIVILIZATION)

CIVILIZATION은 문명/문화와 관련된 용어들로 구성되어 있으며, 축약어로 CV(Civilization)으로 표시한다. 세부 문명/문화로 구분되는 것들만 대상으로 하고, 그 이외의 개체명들은 대상으로 하지 않는다.

6.7.1 CV_NAME

- 정 의
 - 문명/문화 명칭
- 예 제
 - 인더스문명, 거석문화, 날문화
 - 황허문명, 아즈텍문명, 안데스문명, 메소포타미아문명, 도식문화
- 주 의

- 해당 사항 없음

6.7.2 CV_TRIBE

- 정 의
 - 민족/종족 명칭 또한 국가를 구성하는 국민을 지칭
- 예 제
 - 이라크족, 유대인, 조선족, 마오리족, 아메리카인디언, 아리아인, 고산족
- 주 의
 - 해당 사항 없음

6.7.3 CV_SPORTS

- 정 의
 - 스포츠/레포츠/레저 명칭
 - 스포츠/레포츠 종목
- 예 제
 - 스케이팅, 스키, 하이킹, 행글라이딩, 인라인스케이팅, 축구, 마라톤
- 주 의
 - 해당 사항 없음

6.7.4 CV_SPORTS_INST

- 정 의
 - 스포츠/레포츠/레저의 용품/도구 명칭
- 예 제
 - 스케이트, 카빙스키, 행글라이더 등
- 주 의
 - 수영복, 등산복 등의 의복(CV_CLOTHING)의 경우에도 스포츠 용품은 CV_SPORTS_INST으로 태깅한다.

6.7.5 CV_POLICY

- 정 의
 - 제도/정책 명칭
- 예 제
 - 카스트, 쓰레기종량제, 직업공무원제도, 연금제도, 고시제도
 - 가부장제, 진대법, 노예제, 교과담임제, 최저임금제도
 - 교과담임제, 최저임금제도, 섯다운제도
- 주 의
 - 해당 사항 없음

6.7.6 CV_TAX

- 정 의
 - 조세 명칭
- 예 제
 - 소득공제, 면허세, 관세
 - 부가가치세, 근로소득세, 재산세, 취득세, 부가세
- 주 의
 - 해당 사항 없음

6.7.7 CV_FUNDS

- 정 의
 - 연금, 기금, 자금, 펀드 명칭
- 예 제
 - 거치연금, 문예진흥기금, 공적자금, 퇴직연금
 - 노령연금, 개인연금, 통화안정기금, 방송발전기금, 공적자금
- 주 의
 - 해당 사항 없음

6.7.8 CV_LANGUAGE

- 정 의
 - 언어 명칭
- 예 제
 - 힌디어, 프랑스어, 이탈리아어, 산스크리트
 - 라틴어, 몽골어, 아카드어, 말레이어, 스와힐리어
- 주 의
 - 해당 사항 없음

6.7.9 CV_BUILDING_TYPE

- 정 의
 - 건축양식 명칭
- 예 제
 - 기능주의건축, 노르만양식
 - 시토파건축, 에마뮈엘 양식, 그리스건축, 도리스양식
 - 유럽식 목조주택, 기와집, 널벽집, 한옥
- 주 의

- 해당 사항 없음

6.7.10 CV_FOOD

- 정 의
 - 음식/곡물 명칭
 - 음식의 유형도 포함
- 예 제
 - 쇠고기, 소금, 어리굴젓, 명란젓, 국수, 현미, 나물, 해물겨자채, 김치
- 주 의
 - 식물/동물 등의 다른 개체유형과 중복되는 경우에는 다른 개체유형으로 태깅한다.
 - 예) 고등어, 오이, 시금치, 철

6.7.11 CV_DRINK

- 정 의
 - 음료수 명칭
- 예 제
 - 위스키, 우유, 요구르트
 - 주스, 홍차, 레모네이드, 칵테일, 샴페인
- 주 의
 - 해당 사항 없음

6.7.12 CV_CLOTHING

- 정 의
 - 의복/섬유 명칭
 - 의복유형도 포함 (예: 점퍼, 바지, 티셔츠)
- 예 제
 - 모시, 메리야스, 와이셔츠
 - 레이온, 양모, 석면, 점퍼, 카르고팬츠
- 주 의
 - 수영복, 등산복, 등산화 등의 스포츠 용품은 CV_SPORTS_INST으로 태깅한다.

6.7.13 CV_POSITION

- 정 의
 - 직위/직책 명칭
 - 스포츠 포지션
 - 고대 국가의 지위

- 예 제
 - 황제, 학사, 장군, 왕, 영의정, 1품, 대료
 - 사장, 감독자, 교주
 - 4번 타자, 유격수, 투수, 두 번째 투수, 타자
- 주 의
 - 특정 조직의 직위(예: 사장, 이사, 부장, 과장, 대리, 주임, 선임)나 직책(예: 팀장, 실장, 본부장, 사업부장, CEO, CFO, COO)만으로 한정함
 - 조직에서의 위치를 나타내는 말이 아닌 것은 CV_POSITION 대상이 아님
 - 예) 총신, 피서객, 저자, 신, 악신, 선신, 포로, 피서객, 시민권자, 처녀, 아낙네, 유부녀, 재초부, 미혼모, 강습생, 국자학생, 수험생, 연수생, 초심자, 기관투자가, 재일동포, 동포, 대리인, 기업인, 근로자, 국민, 여자, 남자, 하인, 노예

6.7.14 CV_RELATION

- 정 의
 - 인간/가족 관계 명칭
- 예 제
 - 형제, 제자, 남편, 시어머니, 시누이
 - 어머니, 아버지, 친구, 딸
- 주 의
 - 해당 사항 없음

6.7.15 CV_OCCUPATION

- 정 의
 - 직업 명칭
- 예 제
 - 화가, 극작가, 건축가, 가정교사, 기생, 언어학자
 - 공인중개사, 영양사, 군의관, 아나운서, 종군작가
- 주 의
 - 해당 사항 없음

6.7.16 CV_CURRENCY

- 정 의
 - 통화 명칭
- 예 제
 - 달러, 엔, 유로, 루피, 위안, 원
- 주 의

- 해당 사항 없음

6.7.17 CV_PRIZE

- 정 의
 - 상과 훈장
- 예 제
 - 노벨경제학상, 노벨문학상, 노벨물리학상, 노벨생리의학상, 노벨평화상, 노벨화학상, 대통령상, 경제정의기업상, 자기상표제품 수출상, 대중상
- 주의
 - 상과 훈장의 명칭은 DATE, QUANTITY와 함께 쓰인다. 이 경우 DATE와 QUANTITY는 분리하여 태깅한다.
 - 세부 예제는 다음과 같다.

예문) 2000년 제34회 납세자의 날 대통령상

○ <DT_YEAR:2000년> <QT_ORDER:제34회> <DATE:납세자의 날> <CV_PRIZE:대통령상>

✕ <DT_YEAR:2000년> <DATE:제34회 납세자의 날> <CV_PRIZE:대통령상>

✕ <DT_YEAR:2000년> <CV_PRIZE:제34회 납세자의 날 대통령상>

예문) 제10회 경제정의기업상

○ <QT_ORDER:제10회> <CV_PRIZE:경제정의기업상>

✕ <CV_PRIZE:제10회 경제정의기업상>

6.7.18 CV_LAW

- 정 의
 - 법/법률 명칭
 - 법적 권리
 - 범죄 명칭
 - 형벌 명칭
- 예 제
 - 법/법률
 - 국제연합헌장, 민법, 형법, 함무라비법전, 관습법, 헌법, 국회법, 법원조직법
 - 법적 권리
 - 피선거권, 참정권, 선거권, 공민권, 신분권, 친권, 부양청구권, 상속권, 어업권, 상표권
 - 범죄 명칭
 - 간통, 살인죄, 사체유기죄, 폭행죄, 공무집행방해죄, 명예훼손죄, 사기죄
 - 형벌 명칭
 - 화형, 태형, 구금, 구류, 벌금, 몰수, 교수형, 총살형

- 주 의
 - 해당 사항 없음

6.7.19 CV_FOOD_STYLE

- 정 의
 - 음식 종류
- 예 제
 - 한식, 일식, 양식, 중식
 - 보양식, 패스트푸드, 육식, 채식
- 주 의
 - 해당 사항 없음

6.8. 날짜(DATE)

DATE는 날짜 표현을 나타내며, 축약어로 DT(DaTe)로 표시한다.

6.8.1 DT_DURATION

- 정 의
 - 기간
- 예 제
 - 약 5500만~3000만 년 전, 약 40년 동안, BC 624~BC 544년, 10년간, 6~7월
 - ~부터 ~까지, 일년 내내, ~ 간, 전반기/후반기, 성수기/비수기
- 주 의
 - 해당 사항 없음

6.8.2 DT_DAY

- 정 의
 - 특정 일/절기
- 예 제
 - 21일, 20일
 - 입춘, 곡우, 첫날, 첫주, 마지막 말, 내일, 모레, 금일, 당일, 20일 경기
- 주 의
 - 해당 사항 없음

6.8.3 DT_MONTH

- 정 의

- 달
- 예 제
 - 9월, 8월
 - 이달, 저번 달, 올 4월
- 주 의
 - 해당 사항 없음

6.8.4 DT_YEAR

- 정 의
 - 년
- 예 제
 - BC 3000년, BC 300년경, 880년, 1970년대
 - 작년, 내년, 올해, 서기 1910년, 영락 1년
- 주 의
 - '영락'과 같이 연호만 있는 경우는 개체명(DT_YEAR) 대상이 아님

6.8.5 DT_SEASON

- 정 의
 - 계절
- 예 제
 - 봄, 여름, 가을, 겨울
 - 작년 봄, 올 가을, 매년 가을, ~시즌(예: 2008년 시즌, 스토브리그, 포스트시즌)
- 주 의
 - 해당 사항 없음

6.8.6 DT_GEOAGE

- 정 의
 - 지질시대
- 예 제
 - 캄브리아기, 중생대 초기, 중생대, 원생대, 올리고세, 에오세, 신생대 제3기, 신생대, 선사시대, 석탄기, 빙기, 고생대 오르도비스기, 고생대
- 주 의
 - 해당 사항 없음

6.8.7 DT_DYNASTY

- 정 의

- 왕조/시대
- 예 제
 - 청대, 조선시대, 조선 후기, 조선 중기, 명대, 고려시대, 고려 말, 고대 로마제국, 고구려시대, 가리에누스, 황제 시대, 5현제시대, 당대, 고려시대, 조선 정조 23년, 영조 27
- 주 의
 - “구한말”, “일제강점기” 등은 DT_DURATION으로 볼 수도 있지만, 역사의 특정 시대를 나타내므로 DT_DYNASTY로 태깅한다.

6.8.8 DT_OTHERS

- 정 의
 - 기타 날짜
 - DATE의 세부분류에 해당되지 않는 날짜
- 예 제
 - ~년 후/전, ~부터/까지/정도, 지난 세기
 - 제2차 세계대전 전, 임진왜란 때
- 주 의
 - “년/월/일, 년/월, 월/일”이 함께 쓰일 경우에는 전체를 묶어서 DT_OTHERS로 태깅한다.
 - <2007년 10월:DT_OTHERS>에 체급을 올렸다.
 - <작년 11월 25일:DT_OTHERS> OECD에 들어갔다.

6.9. 시간(TIME)

TIME은 시간 표현을 나타내며, 축약어로 TI(Time)으로 표시한다.

6.9.1 TI_DURATION

- 정 의
 - 기간
- 예 제
 - 6시~9시, 오후 1시~3시, 10~15분
 - ~시간 동안
- 주 의
 - 해당 사항 없음

6.9.2 TI_HOUR

- 정 의

- 시간
- 예 제
 - 12시, 저녁 7시
- 주 의
 - 해당 사항 없음

6.9.3 TI_MINUTE

- 정 의
 - 분
- 예 제
 - 56분, 20분
- 주 의
 - 해당 사항 없음

6.9.4 TI_SECOND

- 정 의
 - 초
- 예 제
 - 15초, 0.5초
- 주 의
 - 해당 사항 없음

6.9.5 TI_OTHERS

- 정 의
 - 기타 시간
 - TIME의 세부분류가 아닌 기타 시간
- 예 제
 - 한낮
- 주 의
 - 시/분/초가 연속해서 나오는 경우에는 전체를 TI_OTHERS로 인식함
 - 예) 지금 시간이 2010년 마지막 11시 56분, 57분을 지나고 있습니다.

6.10. 수량(QUANTITY)

QUANTITY는 수량표현을 나타내며, 축약어로 QT(QuanTity)로 표시한다.

6.10.1 QT_AGE

- 정 의
 - 나이
- 예 제
 - 만 55세, 50~60대, 45세, 3살, 18~41세
- 주 의
 - 해당 사항 없음

6.10.2 QT_SIZE

- 정 의
 - 크기/넓이/면적
- 예 제
 - 9,606.33km², 6만 8,000km², 57.7×138.5cm, 430×230 m, 3만 1500km², 30만 정보, 28.02km², 20만 정보, 20km², 2.09km², 2,448km², 2,200km², 19×7.5 m, 192만 2570km², 173만 km²
- 주 의
 - 해당 사항 없음

6.10.3 QT_LENGTH

- 정 의
 - 길이/거리/높이
- 예 제
 - 두께 0.76mm, 가로 54mm, 900m, 8km, 82피트, 7cm, 6~12cm, 3~5cm, 2,200km, 수백 m
- 주 의
 - 해당 사항 없음

6.10.4 QT_COUNT

- 정 의
 - 개수, 빈도
- 예 제
 - 6개, 8백 70선, 하나, 둘
 - 50개국, 1마리, 약 600종, 약 3만 종, 약 1,000종, 9~10마리, 14종, 10권, 두 차례, 9개, 97건, 7천 명~2만 명, 7가지, 77만 9472건, 6가지, 50개국, 5,000자, 4차례, 45만 명, 4,123석, 390여 칸, 2억 1219만 5000명, 18행, 13척, 1,600여 개
- 주 의
 - 해당 사항 없음

6.10.5 QT_MAN_COUNT

- 정 의
 - 인원수
- 예 제
 - 50명, 1명, 약 600명, 12인 등
- 주 의
 - 명확히 인원수를 나타내는 경우에만 태깅한다.

6.10.6 QT_WEIGHT

- 정 의
 - 무게
- 예 제
 - 30kg, 140g, 백근, 9만t, 85~100kg, 60g, 5만 톤, 170t
- 주 의
 - 해당 사항 없음

6.10.7 QT_PERCENTAGE

- 정 의
 - 백분율, 비율, 농도
- 예 제
 - 약 75%, 90%, 8~12ppm, 80~90%, 73.7%, 4배, 400~500ppm, 40%, 3배, 3~5ppm, 32.3%, 30~40%, 2배, 26~34%, 20%, 16%, 11~32%, 10~20ppm, 10ppm, 100ppm, 1/3, 1.1%, 0.719 %, 0.1ppm, 1/4, 4분의 1
- 주 의
 - 해당 사항 없음

6.10.8 QT_SPEED

- 정 의
 - 속도
- 예 제
 - 1km/s, 30km/s
- 주 의
 - 해당 사항 없음

6.10.9 QT_TEMPERATURE

- 정 의
 - 온도
- 예 제
 - 22~25℃, -2.0℃, 9℃~12℃, 41~43.5°
- 주 의
 - 해당 사항 없음

6.10.10 QT_VOLUME

- 정 의
 - 부피
- 예 제
 - 두 말, 800~1,000mm, 500mm, 2만 2000km³, 20ℓ, 2,000~2,500mm, 0.15μg/ℓ
- 주 의
 - 해당 사항 없음

6.10.11 QT_ORDER

- 정 의
 - 순서/순차적 표현
- 예 제
 - 제82호, 제7권, 6기, 65위, 제9회, 제522호, 제4교구, 제2차, 제2대, 제2B족, 제25회, 제1회, 제1대, 제182호, 제12회, 제117호, 제10회, 7층, 5차, 4차, 3층, 25, 1층, 12대
 - 첫, 1위, 재수, 3수
- 주 의
 - 해당 사항 없음

6.10.12 QT_PRICE

- 정 의
 - 금액
- 예 제
 - 72만 원, 60억 달러, 40억 달러, 403원, 290원, 1조 2천억 원, 188원, 100만원, 100만 원
- 주 의
 - 해당 사항 없음

6.10.13 QT_PHONE

- 정 의
 - 전화번호
- 예 제
 - 642-101-1010, (T)<QT_PHONE: 011-111-0111> 등
- 주 의
 - 해당 사항 없음

6.10.14 QT_SPORTS

- 정 의
 - 스포츠 관련 수량 표현
- 예 제
 - 1승2패, 5타점, 3:2, 2연승
 - ~경기, ~승/패/무, 10호골, ~안타
- 주 의
 - 해당 사항 없음

6.10.15 QT_CHANNEL

- 정 의
 - TV/라디오 채널 번호
- 예 제
 - 2채널, 39쇼핑, CJ홈쇼핑
- 주 의
 - 해당 사항 없음

6.10.16 QT_ALBUM

- 정 의
 - 앨범 관련 수량 표현
- 예 제
 - 1집, 2집 등
- 주 의
 - 해당 사항 없음

6.10.17 QT_ZIPCODE

- 정 의
 - 우편번호

- 예 제
 - 110-704, 200-776
- 주 의
 - 해당 사항 없음

6.10.18 QT_OTHERS

- 정 의
 - 기타 수량표현
 - QUANTITY의 세부 분류가 아닌 기타 수량
- 예 제
 - 지식머니 20, 북위 38° 53' 51
- 주 의
 - ‘첫 여행’, ‘첫 날’, ‘첫 주말’, ‘첫 출발’ 등 “첫 + 명사”의 경우에 QT_OTHERS의 대상이 아님

6.11. 사건(EVENT)

EVENT는 특정 사건/사고 명칭을 나타내며, 축약어로 EV(EVent)로 표시한다.

6.11.1 EV_ACTIVITY

- 정 의
 - 사회운동 및 선언
 - 포고문
- 예 제
 - 공민권운동, 노예해방선언, 밀라노칙령, 인권선언, 브나로드운동, 새마을운동, 양무운동
- 주 의
 - "~운동", "~선언"이나 "라노칙령"과 같은 포고문만 해당됨

6.11.2 EV_WAR_REVOLUTION

- 정 의
 - 전쟁/혁명
- 예 제
 - 신해혁명, 포클랜드전쟁, 아편전쟁, 메이지유신
 - 러일전쟁, 청교도혁명, 프롤레타리아혁명, 4.19혁명, 남북전쟁
- 주 의
 - ~전쟁/~혁명만을 대상으로 함

- '메이저유신'은 EV_OTHERS에 해당됨

6.11.3 EV_SPORTS

- 정 의
 - 스포츠 관련 행사
 - 게임 관련 e스포츠 이벤트(행사) 이름
 - 특정 고유행사명이 아니라, 일반 유형(예: 마라톤 대회)의 경우도 포함
- 예 제
 - 올림픽경기대회, 아시아경기대회, 서울 올림픽 경기, 88서울올림픽, 서울올림픽
 - ~대회/경기(예: 거제시장기 전국 윈드서핑대회, 홍경기, 원정경기), ~전(예: 두산전, 개막전, 최종전, 결승전, 예선전, 주중 3연전)
- 주 의
 - "~리그"의 경우는 기관(OGG_ECONOMY)로 볼 수도 있지만 EV_SPORTS로 할당(예: K리그, 메이저리그, 프리미어리그)
 - 스포츠 행사는 보통 개최 회수, 지역 등과 같이 쓰인다.
 - 개최 회수는 분리하여 태깅하고 지역은 통합하여 태깅한다. 세부 예제는 다음과 같다.

예문) 제24회 서울 올림픽 경기

○ <QT_ORDER:제24회> <EV_SPORTS:서울 올림픽 경기>

✕ <QT_ORDER:제24회> <LCP_CAPITALCITY:서울> <EV_SPORTS:올림픽 경기>

✕ <QT_ORDER:제24회 서울 올림픽 경기>

- 스포츠 행사가 DATE와 병행하여 쓰일 경우 명확한 년도 표시는 독립하여 태깅하고, 축약된 경우는 통합하여 태깅한다. 세부 예제는 다음과 같다.

예문) 1988년 서울 올림픽 경기

○ <DT_YEAR:1988년> <EV_SPORTS:서울 올림픽 경기>

○ <EV_SPORTS:88 서울 올림픽 경기>

✕ <EV_SPORTS:1988년 서울 올림픽 경기>

✕ <DT_YEAR:88> <EV_SPORTS:서울 올림픽 경기>

6.11.4 EV_FESTIVAL

- 정 의
 - 축제 명칭
- 예 제
 - 연어축제, 갯방구대축제, 새미골막사발축제, 연등축제, 이천도자기축제
- 주 의
 - "~축제", 영화제만 대상으로 함

6.11.5 EV_OTHERS

- 정 의
 - 기타 사전/사고 명칭
 - ~사태 (ex. 페르시아만사태, 한진중공업파업사태, 중동사태, 천안문사태)
- 예 제
 - 태안 기름유출사고, 북경세계대회
 - 축제 내 체험행사 (예: 바다낚시 체험)
- 주 의
 - 해당 사항 없음

6.12. 동물(ANIMAL)

생물은 2004년 스미스 분류체계에 따라서 다음과 같이 분류된다.

- 진정세균: 세균, 고세균
- 진핵생물
 - 원생동물: 편모충류, 육질충류, 포자충류, 섬모충류
 - 크로미스타: 갈조류, 규조류, 난균류, 착편모조류, 은편모조류
 - 균류: 효모, 곰팡이, 버섯
 - 식물
 - 동물

ANIMAL은 진핵생물 중에서 동물에 해당되는 것으로, 사람을 제외한 짐승을 대상으로 한다. 축약어로 AM(AniMal)로 표시한다. 세분류에 포함되지 않는 동물명은 AM_OTHERS로 태깅한다.

6.12.1 AM_INSECT

- 정 의
 - 곤충
- 예 제
 - 벼메뚜기, 가시거미, 지네, 불나방, 배추흰나비, 참매미, 사마귀
- 주 의
 - 해당 사항 없음

6.12.2 AM_BIRD

- 정 의
 - 새(조류) 이름
- 예 제
 - 왜가리, 뺨꾸기, 거위, 두루미, 황새

- 주 의
 - 해당 사항 없음

6.12.3 AM_FISH

- 정 의
 - 물고기(어류) 이름
- 예 제
 - 참가자미, 전기뱀장어, 연어, 갈치, 멸치, 미꾸리, 송어, 날치
- 주 의
 - 해당 사항 없음

6.12.4 AM_MAMMALIA

- 정 의
 - 포유류
- 예 제
 - 해달, 토끼, 오소리, 돌고래, 고래, 박쥐, 고슴도치, 두더쥐, 당나귀
- 주 의
 - 특정 동물의 이름은 제외 (예: 곰뱀이, 뽀뽀, 누렁이)
 - 사람, 인간, 호모사피엔스 등은 제외

6.12.5 AM_AMPHIBIA

- 정 의
 - 양서류
- 예 제
 - 개구리, 두꺼비, 산개구리, 산청개구리, 도롱뇽, 맹꽁이
- 주 의
 - 해당 사항 없음

6.12.6 AM_REPTILIA

- 정 의
 - 파충류
- 예 제
 - 도마뱀, 거북, 자라, 구렁이, 백사, 공룡, 악어
- 주 의
 - 해당 사항 없음

6.12.7 AM_TYPE

- 정 의
 - 동물 분류 명칭
- 예 제
 - 강장동물, 구두동물, 극피동물, 내항동물, 모악동물, 선형동물, 성구동물, 연체동물, 완보동물, 완족동물, 원생동물, 유선형동물, 유형동물, 윤형동물, 의충동물, 절지동물, 척색동물, 태형동물, 편형동물, 해면동물, 환형동물, 히드로충류, 포유류, 메뚜기과, 어류, 조류, 양서류, 파충류
- 주 의
 - 어류/물고기/민물고기도 AM_TYPE에 해당됨 (AM_FISH, CV_FOOD로 태깅하지 않음)

6.12.8 AM_PART

- 정 의
 - 동물 몸의 한 부분(신체부위) 명칭
 - 일반적으로 눈으로 확인할 수 있는 신체의 겉부분 명칭
- 예 제
 - 팔, 다리, 머리, 장단지, 허벅지
 - 항문, 안구, 이빨
- 주 의
 - 일반적으로 눈으로 확인할 수 있는 겉부분 명칭만 해당되며, '골반' 같은 경우는 TM_CELL_TISSUE에 해당함
 - '껍질'은 물질은 나타내므로, AM_PART에 해당하지 않음

6.12.9 AM_OTHERS

- 정 의
 - 기타 동물 명칭
- 예 제
 - 절지동물(예: 가재), 연체동물(예: 문어, 오징어)
- 주 의
 - "세균", "바이러스", "박테리아", "기생충"은 원래 동물이 아니지만, AM_OTHERS로 할당함

6.13. 식물(PLANT)

PLANT는 식물 명칭을 나타내며, 축약어로 PT(PlanT)로 표시한다. 세분류에 포함되지 않는 식물명은 PT_OTHERS로 태깅한다.

6.13.1 PT_FRUIT

- 정 의
 - 과일 이름
- 예 제
 - 사과, 배, 자두, 바나나, 복숭아, 오렌지, 포도, 호두
- 주 의
 - 해당 사항 없음

6.13.2 PT_FLOWER

- 정 의
 - 꽃 이름
- 예 제
 - 저먼아이리스, 백합, 국화, 제비꽃, 히아신스, 초롱꽃, 할미꽃
- 주 의
 - 해당 사항 없음

6.13.3 PT_TREE

- 정 의
 - 나무 이름
- 예 제
 - 후박나무, 벚나무, 분비나무, 너도밤나무
- 주 의
 - 해당 사항 없음

6.13.4 PT_GRASS

- 정 의
 - 풀 이름
- 예 제
 - 향부자, 하늘타리뿌리, 인삼, 생강, 귀리, 감자
 - 싱아, 갈대, 질경이, 강아지풀, 창포
- 주 의
 - 채소류(예: 양파, 감자, 시금치)도 PT_GRASS에 해당됨

6.13.5 PT_TYPE

- 정 의
 - 식물 유형 명칭
 - 식물의 세부 종을 구분하는 유형만 대상으로 함
- 예 제
 - 녹조류, 선대식물, 양치식물, 종자식물
 - 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
- 주 의
 - 다음은 식물 유형의 대상이 아님
 - 저지대식물, 고산 식물, 꽃, 나무

6.13.6 PT_PART

- 정 의
 - 눈으로 확인 가능한 식물의 한 부분에 대한 명칭
- 예 제
 - 뿌리, 가지, 가시, 결눈, 꽃, 열매, 잎, 줄기, 꽃가루, 꽃송이
- 주 의
 - 눈으로 확인 가능한 부분만 대상으로 함
 - '껍질'은 물질이므로 PT_PART에 해당하지 않음

6.13.7 PT_OTHERS

- 정 의
 - 기타 식물 명칭
 - 식물의 세부분류에 해당하지 않는 기타 식물
- 예 제
 - 우산이끼
- 주 의
 - 해당 사항 없음

6.14. 물질(MATERIAL)

MATERIAL은 물질에 대한 표현을 나타내며, 축약어로 MT(MaTerial)로 표시한다. 물질의 세부 분류에 해당하는 개체명만 대상으로 한다.

6.14.1 MT_ELEMENT

- 정 의
 - 원소명
- 예 제

- 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨
- 주의
 - 원소명은 원소기호와 번호를 할당 받은 물질로 한정한다.
 - 이테르븀 : 주기율표 제3A족에 속하는 희토류원소
 - 크롬 : 주기율표 제6A족에 속하는 크롬족원소의 하나

6.14.2 MT_METAL

- 정 의
 - 금속물
- 예 제
 - 황산철, 알루미늄, 아연, 구리, 수은, 아말감
- 주 의
 - 복합명사에 포함된 금속물일 경우 그 금속물이 복합명사가 나타내는 금속물적인 의미를 표현한다면 금속물로 태깅한다.
 - 예) 마그네슘리본 또는 마그네슘분말로서 카메라 플래시램프·게터·단열재 등에 사용되는 것은 잘 알려진 사실이다

6.14.3 MT_ROCK

- 정 의
 - 암석 명칭
- 예 제
 - 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암
- 주 의
 - 암석 명칭의 경우 ~류 등과 결합하여 암석 종류를 나타낼 때 암석만을 태깅 대상으로 한다.
 - 화강암류 → <MT_ROCK:화강암>류

6.14.4 MT_CHEMICAL

- 정 의
 - 화학물질 명칭
 - 고체 화학물질, 액체 화학물질, 기체 화학물질
 - 단백질/생화학물질/효소/호르몬
- 예 제
 - 고체 화학물질
 - 파라핀, 암모니아, 레이온, 나프탈렌, 수산화나트륨, 염화나트륨, 탄산나트륨, 헤모글로빈, 프리온, 홀로수린, 키틴질, 콜레스테롤, 지방, 에스트로겐, 사포닌, 비타민

- 액체 화학물질
 - 플루오르화수소, 액화석유가스, 과산화수소, 에탄올, 니트로벤젠, 리놀레산,
- 기체 화학물질
 - 황화수소, 이산화탄소, 이산화황, 오존, 일산화탄소, 일산화질소, 네온
- 주 의
 - 해당 사항 없음

6.15. 용어(TERM)

TERM은 위에서 정의된 세부 개체명 이외의 개체명 용어를 나타내며, 축약어로 TM(TerM)으로 표시한다. 아래에 정의된 세부 용어만을 대상으로 하고, 그 이외의 용어는 제외한다.

6.15.1 TM_COLOR

- 정 의
 - 색
- 예 제
 - 흰색, 흰빛, 흑색, 회색, 황, 홍색, 하늘색, 파란빛, 청, 주석백색, 적색, 적갈색, 적, 자주색, 자갈색, 이모색, 은백색, 분홍색, 분백색, 백색, 다갈색, 녹색, 녹갈색, 녹, 노란색, 금색, 검은색, 검은빛, 갈색
- 주 의
 - 해당 사항 없음

6.15.2 TM_DIRECTION

- 정 의
 - 방향을 나타냄
- 예 제
 - 최북단, 중앙부, 중앙, 중부, 서쪽, 서부, 상부, 북쪽, 북서쪽, 북서부, 북부, 북동, 동쪽, 동서, 동북부, 동부, 동남쪽, 동남단, 남쪽, 서쪽, 남서, 남북, 남부, 남동쪽, 남동부, 남동 등
- 주 의
 - 해당 사항 없음

6.15.3 TM_CLIMATE

- 정 의
 - 기후지역 명칭
- 예 제
 - 아열대 기후, 고산기후, 고지기후, 도시기후, 동안기후, 서안기후, 한대기후, 냉대

기후

- 주 의
 - 해당 사항 없음

6.15.4 TM_SHAPE

- 정 의
 - 모양, 형태 명칭
- 예 제
 - 삼각형, 사각형, 팔각형
- 주 의
 - 해당 사항 없음

6.15.5 TM_CELL_TISSUE

- 정 의
 - 세포/조직, 세포 명칭과 생물의 조직 및 기관에 대한 명칭
 - 눈으로 외부적으로 보이지 않는 동물/신체 및 식물의 조직 및 기관 명칭
- 예 제
 - 근육, 극세포, 간상세포, 말세포, 신경세포, 표피, 정원세포, 정모세포
 - 항문, 고막, 식도, 골반, 침샘
- 주 의
 - 신체부위/조직/기관이 특정 단어의 내부에 쓰였을 경우 이를 태깅하지 않는다.
 - 이행부에는 고리 모양의 융기부가 있는데, 이것을 치대(痔帶)·치륜(痔輪)·항문륜(肛門輪)이라고 한다

6.15.6 TMM_DISEASE

- 정 의
 - 증상/증세/질병
- 예 제
 - 기침, 폐결핵, 기관지확장증, 레트증후군, 피고름가래, 호흡불규칙, 호르몬 질환
 - 고산병, 고소공포증
- 주 의
 - '바이러스', '세균'과 같이 질병의 원인은 TMM_DISEASE가 아님

6.15.7 TMM_DISEASE

- 정 의
 - 약/약품명

- 예 제
 - 헤파린, 근이완제, 모기향, 소화제, 아스피린, 구충제
- 주 의
 - 해당 사항 없음

6.15.8 TMI_HW

- 정 의
 - 하드웨어 용어
- 예 제
 - 하드디스크드라이브, 마이크로프로세서, 롬, 디스켓, 자기테이프, 사운드카드, 인공위성
 - 컴퓨터, 모니터, 팩스, 카메라, 디카, 디지털카메라, 냉장고, 전자레인지, 핸드폰, 마우스, MP3 플레이어, 전화(기), 노트북, GPS수신기, TV, dslr 디카, 에어컨, 일반 카드 전화기, 텔레비전, 히터기, GPS, ISE 카드, 네비게이션, 방수 카메라, 충전기
- 주 의
 - TMI_HW도 제품/상품이므로 AF_WARES와 중복되는데, TMI_HW에 우선적으로 할당함
 - TMI_HW는 '카메라'와 같은 제품유형이고, TMI_MODEL는 'D800'과 같은 세부 제품명을 나타냄
 - 예) <초소형 슬라이드 카메라폰:TMI_HW>(<E800:TMI_MODEL>)

6.15.9 TMI_SW

- 정 의
 - 소프트웨어 용어
- 예 제
 - 트로이목마, 인터넷 웜, 웜바이러스, 님다바이러스, 그래픽유저인터페이스, 리눅스, CAD, C언어
- 주 의
 - TMI_SW도 제품/상품이므로 AF_WARES와 중복되는데, TMI_SW에 우선적으로 할당
 - TMI_SW는 '운영체제'와 같은 제품유형이고, TMI_MODEL는 '안드로이드'와 같은 세부 제품명을 나타냄

6.15.10 TMI_SITE

- 정 의
 - url 주소
- 예 제

- <http://www.naver.com>
- 주 의
 - 해당 사항 없음

6.15.11 TMI_EMAIL

- 정 의
 - 이메일 주소
- 예 제
 - aaaa@etri.re.kr
- 주 의
 - 해당 사항 없음

6.15.12 TMI_MODEL

- 정 의
 - 각종 제품 등의 모델명, 제품의 부품류
- 예 제
 - LM1000-2CX
- 주 의
 - TMI_MODEL은 TMI_HW/TMI_SW와 혼동되는 경우가 많은데, 세부 모델명인 경우에만 TMI_MODEL에 해당됨

6.15.13 TMI_SERVICE

- 정 의
 - IT 서비스 용어
- 예 제
 - 와이브로 서비스, DMB 서비스, 위성DMB
 - 카카오톡, 올레닷컴, FTTH, 엑스박스라이브
- 주 의
 - 해당 사항 없음

6.15.14 TMI_PROJECT

- 정 의
 - IT 프로젝트 용어
 - 국가 프로젝트, 회사 프로젝트에 모두 사용
- 예 제
 - BrainKorea21, 4대강사업, 그린IT사업, 무료치과진료사업 등

- 주 의
 - IT 계열에 포함되지만, IT 외의 국가/회사 프로젝트에 모두 사용

6.15.15 TMIG_GENRE

- 정 의
 - 게임 장르
- 예 제
 - RPG, 실시간 전략 게임, 1인칭 슈팅 게임
 - 콘솔 게임, 비디오 게임, PC 게임, 모바일 게임, 온라인 게임
- 주 의
 - 해당 사항 없음

6.15.16 TM_SPORTS

- 정 의
 - 스포츠/레저 용어 (기술, 규칙 이름 등)
- 예 제
 - 밀어내기 득점, 솔로 홈런, 바스켓카운트, 리바운드, 발리슛, 삼진, 볼넷, 사구, 사사구, 1점 홈런, 투런포, 만루홈런, 삼구 삼진, 루킹 삼진
- 주 의
 - 방향성 정보+땅볼/안타/파울/홈런 등은 묶어서 TM_SPORTS로 할당함
 - 예: 1루 강습 안타, 3루 땅볼, 좌전 안타, 3루 펜스쪽 파울, 중월홈런

7. 기타 태깅 원칙

세부분류 개체명 태깅 유의점으로는 최소단위 태깅원칙, 유일 태그 원칙, 최장 세부분류 개체명 태깅 원칙들로 구성되어 있다.

7.1 최소단위 태깅 원칙

2개 이상의 복합 단어의 경우에는 전체를 1개의 개체명으로 태깅하거나, 2개 이상의 단위로 태깅이 가능한 경우가 있다. 이런 경우에는 최소단위로 태깅한다. 세부 예제는 다음과 같다.

예문	태깅 예제
대전 유성 신성동	<LCP_CITY 대전> <LCP_COUNTY 유성> <LCP_COUNTY 신성동>
미국 컬럼비아 대학	<LCP_COUNTRY 미국> <OGG_EDUCATION 컬럼비아 대학>
한국 해양 대학	<OGG_EDUCATION 한국 해양 대학>

대한제국 법무 전라북도재판소	<LCP_COUNTRY 대한제국> <OGG_LAW 법무> <OGG_LAW 전라북도재판소>
미 군정청 광주지방법원	<LCP_COUNTRY 미> <OGG_POLITICS 군정청> <OGG_LAW 광주지방법원>
2000년 제34회 납세자의 날 대통령상	<DT_YEAR 2000년> <QT_ORDER 제34회> <DATE 납세자의 날> <CV_PRIZE대통령상>
제10회 경제정의기업상	<QT_ORDER:제10회> <CV_PRIZE:경제정의기업상>
제24회 서울 올림픽 경기	<QT_ORDER:제24회> <EV_SPORTS:서울 올림픽 경기>
1988년 서울 올림픽 경기	<DT_YEAR:1988년> <EV_SPORTS:서울 올림픽 경기>

7.2 유일 태깅 원칙

개체명 세부 태그 중 일부는 서로 중복 태깅이 가능할 경우가 있다. 세부분류 개체명 태깅은 단일 세부분류 개체명 태깅 원칙을 갖는다. 따라서, 중복태깅 가능한 세부분류 개체명 어휘가 출현했을 경우 우선순위를 정하여 유일태그를 할당해야 한다.

예를 들어, ARTIFACT 개체명의 경우에는 아래와 같이 중복될 수 있으며, 우선순위를 아래와 같이 정한다.

중복유형		예제	태깅 방식
태그1	태그2		
AF_CULTURAL_ASSET	AF_BUILDING	경복궁, 보현사 등	AF_CULTURAL_ASSET에 우선순위
	AF_WORKS	성구백각환 등	AF_WORKS에 우선순위
	AFW_DOCUMENT	신증동국여지승람, 동의보감 등	AFW_DOCUMENT에 우선순위
	AFW_ART_CRAFT	도산도 등	AFW_ART_CRAFT에 우선순위

7.3 최장 세부분류 개체명 태깅 원칙

최장 세부분류 개체명 태깅 원칙은 최장 유형을 태깅 대상으로 하는 것을 말한다.

예를 들면, ‘신고전학과 성장론’은 이론 명칭이 된다. 여기서 ‘신고전학과’라는 FD_SOCIAL_SCIENCE가 포함되어 있으나 최장 세부분류 개체명 태깅 원칙에 따라 ‘신고전학과 성장론’을 TR_SOCIAL_SCIENCE로 태깅한다

7.4 태깅 말뭉치 예제

개체명 태깅 말뭉치는 수작업으로 구축하며, 태깅 말뭉치를 구축하는 포맷 지침은 다음과 같다.

- 문장에 나타난 개체명에 대해서 <>로 범위를 나타낸다.
- <>로 표시된 개체명에 대해서 6장에서 설명된 태그세트 중에서 가장 알맞은 1개 태

그를 <개체명:태그> 형태로 태깅한다.

태깅 말뭉치 구축 예는 다음과 같다.

<아리스토텔레스:PS_NAME>식 <삼단논법:TR_SCIENCE>에서 항과 항의 결합에는 전칭긍정판단, 전칭부정판단, 특칭긍정판단, 특칭부정판단의 <4가지:QT_COUNT>가 있다 .

인식이란, 객관적 실재가 인간의 의식 속에 반영되는 형식의 <하나:QT_COUNT>인데, <아리스토텔레스:PS_NAME>는 이 인식을 체계화시키는 데 필요한 논리적 추리형식으로서 <3단논법:TR_SCIENCE>을 확립하여 <형식논리학:FD_PHILOSOPHY>의 기초를 쌓았다 .

<한문:CV_LANGUAGE>본이 <1784년:DT_YEAR> <가톨릭교회:OGG_RELIGION> 창설을 전후하여 <한국:LCP_COUNTRY>에 들어와 <한글:CV_LANGUAGE>로 번역되어 필사본으로 전해지던 것을 <1892년:DT_YEAR> <위텔:PS_NAME><주교:CV_POSITION>가 전<9책:QT_COUNT>의 활판본으로 간행한 것으로, <한문:CV_LANGUAGE>본에서 <복음성서:AF_WORKS> 해설을 위주로 하는 성경과 잠 부분을 취하고 <프랑스:LCP_COUNTRY> 출신 <예수회:OGG_RELIGION> <선교사:CV_OCCUPATION> <마이야:PS_NAME>가 쓴 동전한문서학서 《<성경광익:AF_WORKS>》에서 의행지덕과 당무지구 부분을 첨가하여 번역한 것이다 .

<한국 IBM:OGG_ECONOMY> 본사는 <서울특별시:LCP_CAPITALCITY> <강남구:LCP_COUNTY> <도곡동:LCP_COUNTY>에 있다 .

7.5 개체명 사용예제

질의응답 기술은 자연어로 질문한 문장을 분석해서 일반 텍스트로부터 정답을 추출해서 제시하는 기술이다. 문서 검색과 정답 추출 기술은 질의응답 기술에서 가장 중요한 기술인데, 이때 개체명 인식이 중요한 역할을 수행한다.

예를 들어, 다음과 같은 질문에 대한 정답을 찾는다고 가정하자.

- “아리스토텔레스가 만든 논리학의 대표적인 이론은?”

위 질문에 대한 질문분석 결과가 다음과 같다고 하자.

- 질의어(질문 키워드): ‘아리스토텔레스’, ‘만들다’, ‘논리학’, ‘대표’, ‘이론’
- 질문초점: 이론
- 정답유형: TR_SCIENCE

위 질문에 대해서 정답을 찾기 위해서는 질문분석 결과 중 질의어과 관련된 문서/문장을 검색하고, 검색된 문서/문장으로부터 질문에 가장 적합한 정답을 추출해야 한다.

문서/문장 검색으로 다음과 같은 질문과 가장 유사한 문장들이 검색되었다고 하자.

아리스토텔레스식 삼단논법에서 항과 항의 결합에는 전칭긍정판단, 전칭부정판단, 특칭긍정판단, 특칭부정판단의 4가지가 있다.

인식이란, 객관적 실재가 인간의 의식 속에 반영되는 형식의 하나인데, 아리스토텔레스는 이 인식을 체계화시키는 데 필요한 논리적 추리형식으로서 3단논법을 확립하여 형식논리학의 기초를 쌓았다.

검색된 위의 문장에서 정답을 추출하기 위해서는 질문초점인 ‘이론’과 관련된 개체명 유형인 ‘TR_SCIENCE’를 정답유형으로 설정하고, 문서/문장 검색 결과로부터 ‘TR_SCIENCE’인 개체명들을 정답후보로 추출해서 질문과 가장 적합한 1개를 정답으로 제시하게 된다.

위의 검색 문장에 대한 개체명 인식 결과는 다음과 같다.

<아리스토텔레스:PS_NAME>식 <삼단논법:TR_SCIENCE>에서 항과 항의 결합에는 전칭긍정판단, 전칭부정판단, 특칭긍정판단, 특칭부정판단의 <4가지:QT_COUNT>가 있다 .

인식이란, 객관적 실재가 인간의 의식 속에 반영되는 형식의 <하나:QT_COUNT>인데, <아리스토텔레스:PS_NAME>는 이 인식을 체계화시키는 데 필요한 논리적 추리형식으로서 <3단논법:TR_SCIENCE>을 확립하여 <형식논리학:FD_PHILOSOPHY>의 기초를 쌓았다 .

마지막으로 검색 문장의 개체명인식 결과로부터 질문분석 결과 중 정답유형 ‘TR_SCIENCE’에 해당하는 개체명인 ‘삼단논법’과 ‘3단논법’을 정답 후보로 제시한다.

<아리스토텔레스:PS_NAME>식 <삼단논법:TR_SCIENCE>에서 항과 항의 결합에는 전칭긍정판단, 전칭부정판단, 특칭긍정판단, 특칭부정판단의 <4가지:QT_COUNT>가 있다 .

인식이란, 객관적 실재가 인간의 의식 속에 반영되는 형식의 <하나:QT_COUNT>인데, <아리스토텔레스:PS_NAME>는 이 인식을 체계화시키는 데 필요한 논리적 추리형식으로서 <3단논법:TR_SCIENCE>을 확립하여 <형식논리학:FD_PHILOSOPHY>의 기초를 쌓았다 .

표준작성 공헌자

표준 번호 : TTAx.xx-xx.xxxx/R1

이 표준의 제.개정 및 발간을 위해 아래와 같이 여러분들이 공헌하셨습니다.

구분	성명	위원회 및 직위	연락처	소속사
표준(과제) 제안	최미란	메타데이터 프로젝트그룹 위원	miranc@etri.re.kr	ETRI
표준 초안 작성자	최미란	메타데이터 프로젝트그룹 위원	miranc@etri.re.kr	ETRI
	이충희	-	forever@etri.re.kr	ETRI
	임준호	-	joonho.lim@etri.re.kr	ETRI
	임수중	-	isj@etri.re.kr	ETRI
	김현기	-	hkk@etri.re.kr	ETRI
표준 초안 에디터	김장원	메타데이터 프로젝트그룹 위원	ikaros1223@gmail.com	KISTI
표준 초안 검토	서태설	메타데이터 프로젝트그룹 의장	tsseo@kisti.re.kr	KISTI
		외 프로젝트그룹 위원		
표준안 심의	박승민	소프트웨어/콘텐츠 기술위원회 의장	minpark@etri.re.kr	ETRI
		외 기술위원회 위원		
사무국 담당	김영화	부장	ykim@tta.or.kr	TTA
	이상호	선임	sangho.lee@tta.or.kr	TTA

정보통신단체표준(국문표준)

개체명 태그세트 및 태깅 말뭉치
(Tag Set and Tagged Corpus for Named Entity Recognition)

발행인 : 한국정보통신기술협회 회장

발행처 : 한국정보통신기술협회

463-824, 경기도 성남시 분당구 분당로 47

Tel : 031-724-0114, Fax : 031-724-0019

발행일 : 2015.xx
