

국립국어원 2020-01-21

발 간 등 록 번 호
11-1371028-000834-01

## 2020년 개체명 말뭉치 연구 분석

사업 책임자  
차 정 원





# 제 출 문

국립국어원장 귀하

국립국어원과 체결한 용역 계약에 따라 ‘2020년 개체명 말뭉치 연구 분석’에 관한 용역 보고서를 작성하여 제출합니다.

■ 사업기간: 2020년 06월 ~ 2020년 12월

2020년 12월 22일

사업 책임자: 차 정 원 (데이터리)

사업 수행 기관 (주)데이터리, 한림대학교

사업 책임자 차정원

공동 연구원 신서인, 이선웅

사업 참여자 오진영, 이수현, 전상호, 이한범  
박진현, 임윤정, 황지영, 임윤정  
박경희, 박은실, 배진영, 최윤진



<사업 수행자>      데이터리, 한림대학교

사업 책임자	차정원 (데이터리 대표)
사업 참여자	신서인 (한림대학교 교수)
	이선웅 (경희대학교 교수)
	오진영 (데이터리 이사)
	이수현 (한림대학교 시간강사)
	이한범 (한림대학교 시간강사)
	박진현 (한림대학교 박사수료)
	전상호 (한림대학교 박사과정)
	임윤정 (경희대학교 시간강사)
	황지영 (경희대학교 시간강사)
	박경희 (경희대학교 시간강사)
	박은실 (데이터리 전임연구원)
	배진영 (데이터리 전임연구원)
	최윤진 (데이터리 전임연구원)



## 2020년 개체명 말뭉치 연구 분석

전 세계적으로 4차 산업혁명이 도래하고 인터넷 기술과 인공지능, 빅데이터 등의 디지털 기술이 나날이 발전하고 있는 지금, 인터넷 매체는 가장 중요한 언어 소통 환경으로 자리잡았다. SNS 및 커뮤니티 등의 발전으로 자신의 생각과 경험을 전달하고 타인과 의사소통을 하는 데에 인터넷 매체를 이용하는 일이 점점 더 늘어나고 있다. 다시 말해서 인터넷 속 언어의 양은 하루가 다르게 방대해져 가며 이를 분석하여 데이터화 하는 작업의 가치는 점점 중요해져 간다는 뜻이다.

이러한 시대의 흐름을 반영하여 ‘2020년 개체명 말뭉치 연구 분석’ 사업은 ‘19년 국립국어원 구축 웹 원시 말뭉치’를 대상으로 4차 산업혁명에 대비한 인공지능 언어 처리 기술 개발 및 국어 연구의 기반으로 활용할 수 있는 대규모 고품질 개체명 분석 웹 말뭉치 구축을 목표로 진행되었다. 이와 함께 ‘개체명 태그 세트 및 태깅 말뭉치’ (한국정보통신기술협회, 2015) 및 ‘세부분류 개체명 가이드라인 2018’ (한국전자통신연구원 언어지능연구그룹 SW콘텐츠연구소, 2018. 12.)의 146개 세부분류 체계를 기준으로 하여 작성된 ‘19년 국립국어원 개체명 분석 말뭉치 지침’을 재정비하고 세밀화하는 작업도 함께 진행하여 앞으로 진행될 개체명 분석 말뭉치 구축 작업의 기준이 될 수 있도록 개체명 말뭉치 분석 지침을 수정하고 보완하였다.

사업 기간은 약 7개월이며 사업 기관인 데이터리와 공동사업기관인 한림대학교는 국립국어원으로부터 블로그와 웹 게시판 등에서 수집한 ‘19년 국립국어원 구축 웹 원시 말뭉치 500만 어절’을 전달받아 작업을 진행하였다. 전체적인 개체명 구축 프로세스 및 구축 시스템과 도구 개발 그리고 구축 작업자 관리 및 교육은 데이터리에서 맡았으며, 국어학적 지식을 기반으로 하여 기존의 지침을 수정 및 보완하는 작업과 말뭉치를 검증하기 위한 평가 말뭉치 구축 작업은 한림대학교가 맡아 진행하였다. 효율적인 방식으로 성공적인 결과를 끌어내기 위하여 본 사업은 다음과 같이 진행되었다.

첫째, 데이터리에서 보유하고 있는 개체명 구축 프로세스와 도구를 이용하였다.

일관성 검사를 기반으로 하는 레이블링·검증 도구와 특허 인증 기술인 Rough set(집합에서 각 항목을 구성하는 속성들 간의 비교를 통해 분류 경계 영역을 계산하는 방법)을 이용한 오류 판별 방법, RDR(Ripple-Down Rule)을 기반으로 하는 오류 자동 수정 방법 등 자동화 시스템을 통하여 작업을 수행하였다. 이 방법은 분류를 결정하는 하나의 일반적인 규칙이 존재할 때, 그 규칙으로부터 예외 조건이 발생하는 각 경우를 명시한다. 예외 조건들 중에 오류일 가능성이 높은 항목을 선별하고 조건에 해당하는 문장을 수정함으로써 오류를 줄여나가는 방법이다. 이것을 한국전자통신연구원(ETRI)

지침에서 분류한 146개(최종 지침 수정 후 150개) 개체명 범주를 자동화 시스템에 적용한 후, 개체명 말뭉치의 구축·정제·오류 검출 작업을 진행하였다. 자동화 시스템의 작업 과정은 오류 검출부터 수정까지 전문가의 관리하에 일괄적으로 이루어지므로 작업의 신속성과 정확성을 모두 얻을 수 있는 방식이다. 이 작업을 통해 산출된 결과물을 전문가가 구축한 평가 말뭉치와 비교하여 결괏값을 산출하고 오류를 수정하였다. 또한, 작업자에 의해 구축된 모든 결과물은 자동화 시스템의 성능 향상을 위한 자료로 다시 이용되었다. 작업 중 발생한 오류-수정 쌍을 이용하여 오류 수정 규칙을 자동으로 생성하고 시스템 내 전체 말뭉치에 적용하여 일괄 수정 작업을 거쳐 최종적으로 말뭉치의 품질을 향상시키는 방식을 사용하였다.

둘째, 클라우드 환경을 지원하는 말뭉치 구축 시스템과 개체명 구축 전문 도구 ‘kane’을 이용하였다.

다수의 사용자가 클라우드 환경에서 동시에 작업을 진행하고 구축된 데이터를 즉각적으로 전달함으로써 작업이 빠르게 진행될 수 있었다. 또한 다수의 전문가에 의한 다면 평가가 이루어져 결과 말뭉치의 품질을 높이고 전문가의 정성적 검토가 이루어짐으로써 정량적인 평가의 단점 또한 보완할 수 있었다.

또한, 말뭉치를 구축하는 도구로는 데이터리에서 보유하고 있는 프로그램 ‘kane’을 사용함으로써 작업의 효율성을 높이고 단순 실수를 낮췄다. ‘kane’은 개체명 구축 작업에 맞게 개발된 도구로 내부적인 기능과 구축물 업로드, 다운로드, 사전 찾기, 우리말샘과의 연동 등 개체명 구축 작업에 특화된 도구이다. 작업자들은 자동화 시스템을 통해 전처리된 말뭉치를 ‘kane’을 이용하여 구축 작업을 진행하였다.

마지막으로 한림대학교와 경희대학교 소속의 국어학 전문가들이 다수 참여하였다.

관련 국어 연구와 사업에 참여한 경험이 있는 전문가들은 기존에 작성된 지침에 대한 문제점과 수정 방안을 제시하고 국립국어원과 의논하여 도출된 의견을 지침에 반영하였다. 또한, 평가 말뭉치를 구축함으로써 전문가 집단의 결과와 작업자 집단의 결과를 비교하여 수정할 수 있도록 하였고 이를 통해 결과물에 대한 오류율을 줄이고 말뭉치의 완성도를 높이는 데 일조하였다.

언어 소통 환경에서 인터넷이 차지하는 비중이 점점 커지고 있는 상황이므로 웹 언어 자료는 언어 연구와 언어 처리 산업에서 활용도가 커질 수밖에 없다. 하지만 높은 활용도에 비해 웹 언어 개체명 분석 자료의 양은 현저히 부족하다. 본 사업은 이러한 현 상황을 인지하고 다방면에서 활용할 수 있는 개체명 분석 웹 말뭉치 구축을 위해 진행되었다. 데이터리와 한림대학교가 참여한 ‘2020년 개체명 말뭉치 연구 분석’ 사업의 결과물은 국어 연구와 인공지능 및 빅데이터 산업에 도움이 될 수 있도록 공공재로 사용될 것이며 나아가 4차 산업혁명으로 나아갈 수 있는 디딤돌이 될 수 있을 것이다.

**주요어:** 웹 원시 말뭉치, 개체명, 인터넷 언어, 지침, 검증, 구축





# 차 례

## 제1장 사업 개요

1. 사업 목적 및 범위 .....	3
2. 사업 수행 방법 개관 .....	4
3. 사업 수행 절차 .....	5
4. 사업 추진 경과 .....	6

## 제2장 사업 수행 내용

1. 사업 수행 방식 .....	9
2. 개체명 분석 말뭉치 구축 도구 .....	11
3. 개체명 구축 작업 과정 .....	17
4. 구축 말뭉치 품질 검증 .....	19
5. 개체명 분석 말뭉치 구축 인력 .....	21

## 제3장 개체명 분석 말뭉치 구축 지침

1. 지침 주요 수정사항 .....	33
2. 개체명 분석의 대원칙 정비 .....	38
3. 세부분류별 분석 지침 정비 .....	39

## 제4장 사업 수행 결과

1. 사업 요약 .....	51
2. 최종 산출물 .....	52
3. 남은 문제 .....	57

부록 1. 2020년 개체명 분석 말뭉치 구축 최종 지침

참고 문헌

<Abstract>

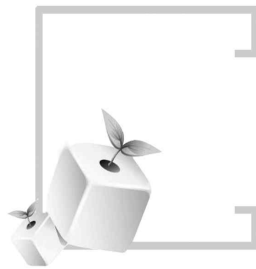
## 표 차례

<표 1> 사업 추진 경과 .....	6
<표 2> 사업 진행 과정 .....	9
<표 3> 작업 준비단계 회의 진행 내용 .....	9
<표 4> ‘꽃’ 클러스터 문장 표 .....	12
<표 5> 말뭉치 구축 과정 .....	17
<표 6> 개체명 부착 작업자 작업 방식 .....	18
<표 7> 검토/평가 작업자 작업 방식 .....	18
<표 8> 말뭉치 구축 작업 내용 .....	18
<표 9> 품질 측정 결과 .....	21
<표 10> 2020년 개체명 말뭉치 연구 분석 사업 참여자 명단 .....	23
<표 11> 교육 진행 내역 .....	24
<표 12> 세미나 개최 내역 .....	27
<표 13> 지침 주요 수정사항 .....	33
<표 14> 지침 수정 관련 회의 내용 .....	43
<표 15> 최종 납품물 .....	52
<표 16> 원시 말뭉치 JSON 형식 .....	53
<표 17> 상위 15개 카테고리 출현 빈도 .....	53
<표 18> 세부분류 전체 빈도 .....	55

# 그림 차례

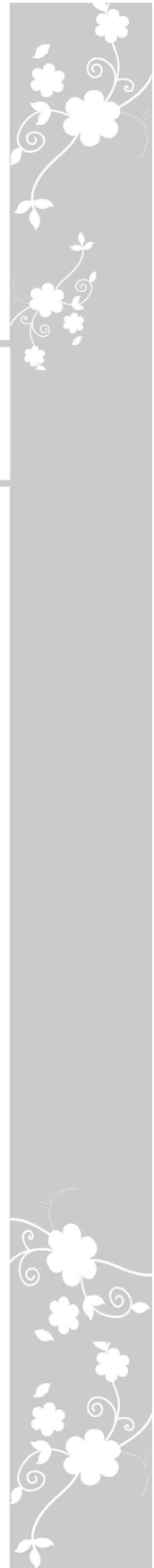
<그림 1> 전체 사업 수행 절차 .....	5
<그림 2> 구축 도구 kane 사용 방법 .....	13
<그림 3> 구축 도구 kane 사용 방법 2 .....	14
<그림 4> kane 툴 개체명 포맷 및 출력 화면 사용 방법 .....	14
<그림 5> kane의 세부 기능 .....	16
<그림 6> 사업 참여 기관 .....	19
<그림 7> 사업 참여 관련 작업자 제출 문서 .....	25
<그림 8> 작업자 질문사항 .....	26
<그림 9> 개체명 작업 제출 현황 .....	27
<그림 10> 구글 시트 .....	29
<그림 11> 구글 시트 활용 방법 .....	29





## 제 1 장

# 사업 개요





# 제1장 사업 개요

## 1 사업 목적 및 범위

‘2020년 개체명 말뭉치 연구 분석’ 사업은 4차 산업혁명 시대에 들어서면서 언어처리 활용도 측면에서 중요성이 높아지고 있는 웹 환경의 언어 자료에 대한 가치와 특성을 인지하고, 이를 개체명 분석 연구와 결합하여 국립국어원에서 2019년에 구축한 웹 원시 말뭉치에 개체명을 부착하는 사업이다. 본 사업에서는 활용도가 매우 높으나 관련 데이터가 부족한 개체명 분석 웹 말뭉치를 구축함과 동시에 구축에 필요한 개체명 지침을 명확하게 정립하여 이후에 이어질 연구의 방향성과 분석의 기준을 제시하고자 한다. 본 사업을 통해 구축된 개체명 분석 웹 말뭉치는 인공지능 언어처리 기술 개발 및 국어 연구의 기반 자료로 활용할 수 있도록 배포되어 이를 활용한 산업과 연구의 발전에 도움을 주는 데 목적을 두고 있다. 본 사업의 수행 범위는 다음과 같다.

첫째, 국립국어원에서 2019년에 구축한 웹 원시 말뭉치 500만 어절을 대상으로 개체명 세부분류 150개<sup>1)</sup>에 속하는 개체명을 분석하고 표지를 부착한다.

분류 항목의 수가 많은 만큼 반복된 검증을 통해 오류율을 최소화하고, 데이터가 보유하고 있는 구축 프로세스 및 시스템을 적극 활용하여 분석 정확률 95% 이상의 고품질 개체명 분석 말뭉치 구축을 목표로 한다.

둘째, 개체명 분석 말뭉치 구축 작업을 위한 지침을 정리한다.

150개의 세부분류 개체명 분석을 위해서는 명확한 기준과 정의가 기술된 지침이 필요하다. 이를 위해 ‘개체명 태그 세트 및 태깅 말뭉치’<sup>2)</sup> 및 ‘세부분류 개체명 가이드라인 2018’<sup>3)</sup>을 기반으로 하여 작성된 ‘19년 국립국어원 개체명 분석 말뭉치 지침’을 수정·보완하고 세밀화하여 향후 관련 사업에서도 본 사업에서 정립된 지침이 기준으로 사용될 수 있도록 한다.

마지막으로, 사업 종료 후 납품된 자료의 품질 보증 및 보완 체계를 수립한다.

납품된 자료에 수정이 필요하거나 오류가 있을 경우, 하자 보수를 위한 체계를 설계하고 국어원의 요구사항 및 의견 등을 즉각적으로 반영할 방법을 설계하여 빠르고 신속하게 결과물의 보완이 이루어질 수 있도록 한다.

1) 최말단 태그의 총 개수임.

2) 한국정보통신기술협회, 2015.

3) 한국전자통신연구원 언어지능연구그룹 SW콘텐츠연구소, 2018.12.



본 사업팀은 사업 진행 과정 전반을 위 사업 범위에 맞추어 설정하였고 ‘2020년 개체명 말뭉치 연구 분석’ 사업이 종료 기간 내에 성공적으로 마무리 될 수 있도록 진행하였다.

## 2 사업 수행 방법 개관

본 사업의 참여 기관은 데이터리(Data.ly)와 한림대학교 두 기관이다. 대량의 고품질 개체명 분석 웹 말뭉치 구축과 ‘19년 국립국어원 개체명 분석 말뭉치 지침’을 수정 및 보완하는 작업을 효율적으로 수행하기 위해 각 기관의 전문성을 발휘할 수 있는 과업을 중심으로 말뭉치 구축팀(이하 구축팀)’과 ‘품질 검증팀(이하 검증팀)’으로 분업하여 사업을 진행하였다.

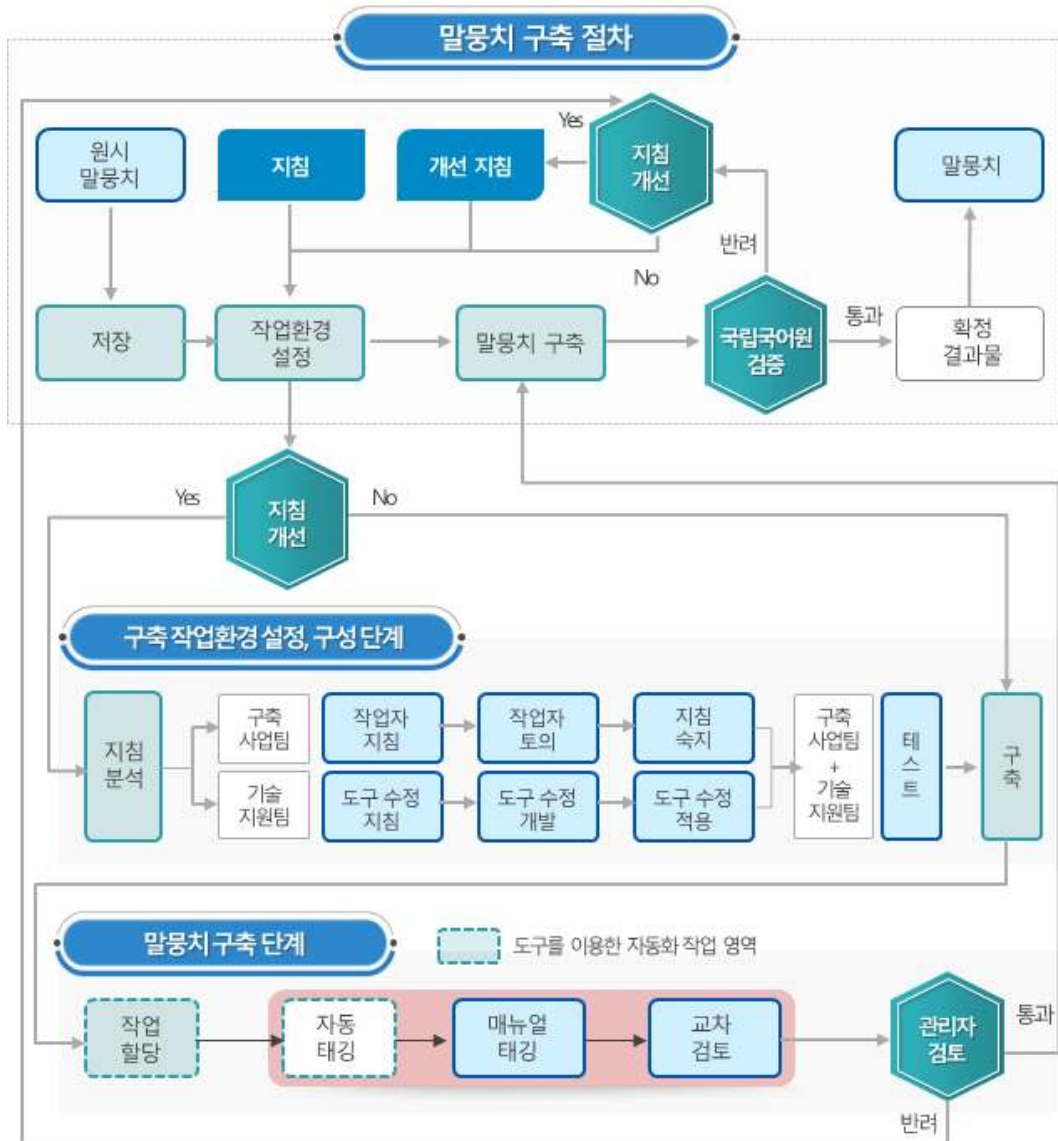
데이터리는 사업의 전반을 총괄하며 ‘구축팀’으로서의 역할을 수행하였다. 데이터리가 자체적으로 보유한 개체명 자동 분석 시스템과 프로세스를 이용하여 1차적으로 개체명 분석 말뭉치를 구축한다. 이를 정제된 말뭉치로 구축하기 위해 작업자를 모집하여 교육을 진행한 다음 개체명 구축 도구 ‘kane’을 이용하여 정제 작업을 진행하였다. 또한, 자동화 시스템을 이용하여 구축된 말뭉치의 오류 검출 및 수정 작업을 진행하였다.

한림대학교는 국어학적 지식을 지닌 전문가가 중심이 되어 ‘검증팀’의 역할을 수행하였다. 작업자들이 구축한 말뭉치에 대한 검증 및 관리를 위하여 전문가들이 직접 전체 웹 말뭉치 중 약 10%를 구축하였고 완성된 평가 말뭉치는 자동화 시스템에 반영되어 오류 검출과 수정 작업에 이용되었다. 또한, 동시에 ‘19년 국립국어원 개체명 분석 말뭉치 지침’의 검토를 진행하였다. 국어학 전문가와 빅데이터 전문가의 자문을 구하여 지침에서 추가 및 삭제되어야 할 부분을 확인하고 국립국어원과 협의하여 최종적으로 도출된 결과를 지침에 반영하는 과정을 통해 지침을 수정, 보완하였다.

구축팀과 검증팀의 말뭉치 구축 작업 참여자들은 모두 국립국어원 말뭉치 구축 지원 시스템에 접속하여 국립국어원에서 제시하는 말뭉치를 대상으로 개체명 분석 작업을 실시하였으며 이 결과는 작업자들의 지침 이해도 평가, 지침 보완 등에 활용되었다.

### 3 사업 수행 절차

전체적인 사업 진행 절차는 아래 그림과 같다.



<그림 1> 전체 사업 수행 절차

우선 구축 작업이 진행되기 전에 ‘19년 국립국어원 개체명 분석 말뭉치 지침’을 검토, 수정한 후 1차 지침이 확정되면 말뭉치 구축 작업을 진행한다. 먼저 자동 분석 시스템을 이용하여 자동으로 태깅이 가능한 부분을 구축하고 나머지 부분은 지침 교육을 이수한 작업자들이 투입되어 작업을 진행한다. 표지가 부착된 작업물은 자동화 시스템을 통하여 교차 검토되고, 여기서 발견된 오류들

을 관리자가 검토하여 수정하고 필요한 부분은 지침에 반영한다. 이러한 과정이 수차례 반복되면서 작업물은 구축과 수정을 반복하게 되고, 지침 또한 수정을 거듭하게 되어 최종적으로 완성된 개체명 분석 말뭉치를 국립국어원이 검증하게 된다. 국립국어원의 수정 요구사항까지 반영되어 승인을 받은 말뭉치와 지침은 ‘2020년 개체명 말뭉치 연구 분석’ 사업의 최종 결과물로 제출된다.

절차 진행 단계별 자세한 설명은 이후 ‘2-1 사업 수행 방식’에서 다루도록 한다.

## 4 사업 추진 경과

본 사업의 추진 경과는 다음과 같다.

사업 추진 경과								
일련 번호	추진 내용	월						
		6	7	8	9	10	11	12
1	<ul style="list-style-type: none"> <li>■ 국립국어원 구축 지원</li> <li>■ 시스템 사전 작업</li> <li>■ 말뭉치 지침 분석</li> <li>■ 말뭉치 검증 시스템 준비</li> </ul>							
2	<ul style="list-style-type: none"> <li>■ 말뭉치 구축</li> <li>■ 오류 탐지</li> <li>■ 오류 수정</li> </ul>							
3	<ul style="list-style-type: none"> <li>■ 전문가 평가</li> <li>■ 통계적 평가</li> </ul>							
4	■ 말뭉치 구축 마무리							
5	<ul style="list-style-type: none"> <li>■ [산출물]</li> </ul> 개체명 분석 지침							
6	<ul style="list-style-type: none"> <li>■ [산출물]</li> </ul> 최종 보고서							

<표 1> 사업 추진 경과



## 제 2 장

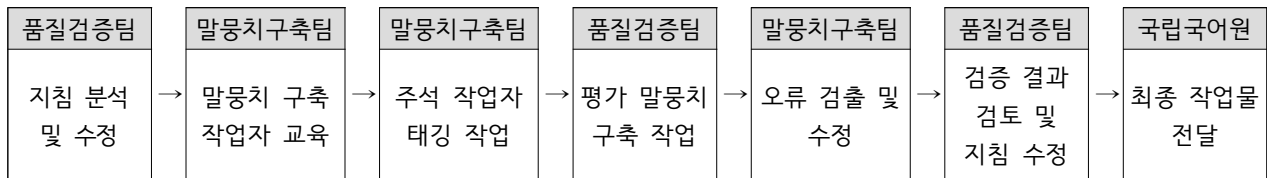
# 사업 수행 내용





## 제2장 사업 수행 내용

### 1 사업 수행 방식



<표 2> 사업 진행 과정

성공적인 사업 진행을 위하여 구축팀과 검증팀의 역할에 맞게 과정별로 업무를 분담하였다. 사업의 전반적인 진행 과정은 위 표 2와 같다. 진행 과정은 크게 작업 준비 단계, 구축 단계, 검증 단계, 마무리 단계 총 4단계로 나눌 수 있으며 단계별 자세한 진행 내용은 다음과 같다.

작업 준비 단계에서는 사업을 원활하게 진행하기 위해 사업 수행 기관인 데이터리(Data.ly)와 한림대학교 간 협의를 진행하였다. 사업의 목표와 내용, 일정을 공유하고 각 기관이 맡은 업무 내용을 명확히 하였다. 또한 개체명 분석의 기준이 될 ‘19년 국립국어원 개체명 분석 말뭉치 지침’에서 수정이 필요한 항목에 대해 논의하였다.

일시	차수	참여 기관	내용
20.06.19	1차	데이터리 한림대학교	- 사업 목표 및 작업 내용 숙지 - 진행 일정 공유 - 19년 지침 항목별 문제 검토 및 토의
20.07.06	2차	데이터리 한림대학교	- 19년 지침 항목별 문제 검토 및 토의
20.07.13	3차	데이터리 한림대학교	- 19년 지침 항목별 문제 검토 및 토의

<표 3> 작업 준비 단계 회의 진행 내용

구축팀에서는 말뭉치 구축 작업 시작을 위한 전처리, 구축, 평가 검증 시스템을 정비하여 작업이 바로 진행될 수 있도록 준비하였다. 국립국어원에서 전달받은 ‘웹 원시 말뭉치’를 1차 정제하여 개체명 자동 분석 시스템으로 구축이 가

능한 부분을 파악하고 작업자들에게 즉시 문서 분배가 가능하도록 준비하였다. 1차 정제는 다음 장에 설명할 도구 중 전처리 도구를 이용한 것으로 구축 도구(kane)에서 사용할 수 있는 형태 변경과 군집화가 진행된다. 또한, 구축 작업에 바로 투입될 수 있도록 구축 작업자들을 교육하였다. 작업자 교육에서는 지침에 대한 안내와 구축 작업 실습이 이루어졌다. 자세한 내용은 ‘2-3 말뭉치 구축 인력’에 상세히 기술되어 있다.

다음은 구축 단계이다. 구축 단계에서는 구축팀에서 보유하고 있는 자료로 학습한 시스템을 이용하여 초기 말뭉치를 구축하였다. 초기 말뭉치는 준비 단계에서 얻어진 1차 정제에 자동화 시스템 결과를 추가한 것을 말한다. 자동화 시스템은 인공지능(AI) 언어 분석 기술을 이용한 모델로, 학습한 자료를 바탕으로 원시 말뭉치에 자동으로 개체명을 빠르게 부착할 수 있다. 초기 말뭉치는 검증 단계에서 작업자의 결과와 비교하여 오류 후보를 파악하는 데 사용할 수 있으며, 구축 단계 전 대량의 문장 중 우선적으로 작업할 문장을 선택하여 작업 분배 일정을 고려하는 검토 요소가 된다. 초기 말뭉치를 개체명 부착 작업자들에게 전달하면 지침을 숙지한 작업자들은 프로그램이 잘못 부착한 개체명과 부착하지 않은 부분을 데이터리가 개발한 개체명 구축 도구인 ‘kane’을 이용하여 수정하고 부착하였다. 이와 동시에 품질 검증팀에서는 검증 단계에서 사용될 평가 말뭉치를 구축하였다.

구축 작업이 어느 정도 진행되면 검증 단계가 동시에 진행된다. 검증 단계에서는 검증팀에서 구축한 평가 말뭉치를 이용하여 다면평가를 진행하였다. 평가 말뭉치를 반영한 자동 오류 탐지 시스템을 통해 작업자들의 구축 결과물에서 오류를 검출하고 이를 작업자가 수정하도록 피드백을 실행하였다. 위 과정은 계속 반복하여 진행되며 도출된 오류 사항들은 검증팀으로 전달되어 지침의 재정리 및 정비에 참고 자료로 쓰이기도 하고, 자동화 시스템에 적용되어 성능을 높이는 데 쓰였다. 오류에 대해 새로 학습한 시스템은 문서에 재적용되며 작업의 정확성과 일관성을 높이게 된다.

마지막으로 종료 단계에서는 작업한 말뭉치 구축 결과를 정리하고 이를 분석한 통계 정보를 도출해 내었다. 말뭉치 품질을 확인한 뒤 검출된 오류를 수정하고 지침 변경 사항들을 다시 말뭉치에 적용하는 작업을 반복하여 목표치인 95%에 도달할 수 있도록 하였다. 지침 또한 여러 차례의 수정을 거쳐 국립국어원의 승인을 받아 최종적으로 검수를 하였다.

전 단계를 거쳐 완성된 결과물인 ‘개체명 분석 웹 원시 말뭉치 500만 어절’과 ‘20년 국립국어원 개체명 분석 말뭉치 지침’을 국립국어원에 제출하였다.

제출한 결과물은 국립국어원 자체 검증팀에서 검증 작업이 이루어지며 검증이 끝난 뒤 결과는 구축팀과 검증팀에게 전달된다. 마지막 검증 과정에서 발생한 오류나 지침 변경으로 인해 수정되어야 하는 부분들을 최종 보완하고 사업에 대한 최종 보고를 마치면 ‘2020년 개체명 말뭉치 연구 분석’ 사업이 종료된다.

## 2 개체명 분석 말뭉치 구축 도구

### □ 사업 추진 전략 및 지원 체계

본 사업 목표를 달성하기 위하여 구축팀과 검증팀이 유기적으로 작업할 수 있도록 지원이 필요하다. 구축팀은 지침에 맞게 말뭉치를 구축하는 것이 목표이고 검증팀은 지침의 보완, 수정, 구축 및 구축팀의 구축 결과물에서에서 수정이 필요한 항목과 누락된 항목 등을 전달하여 말뭉치의 품질을 향상시키는 역할을 한다. 검증팀과 구축팀의 협업을 지원하기 위해 고품질 말뭉치 구축을 위한 도구를 제공하고 먼저 학습된 전문가의 지식을 통해 구축팀의 작업을 보조하며, 검증팀의 피드백을 반영하는 역할을 한다.

### □ 전처리 도구

전처리를 통하여 개체명 말뭉치 구축의 효율성을 높이고 일관성 확인 등의 검수를 수행할 수 있다. 전처리 도구의 첫 번째 기능은 말뭉치를 구축 도구(kane)에서 사용할 수 있는 형태로 변경하는 것이다. JSON 포맷의 대상 원시 말뭉치를 구축 도구에서 사용할 문장으로 분리한다.

두 번째는 유사한 개체명이 있는 문장들을 군집화(Cluster, 클러스터)하는 기능이다. 군집화를 통하여 유사한 개체명을 포함한 문장들이 하나의 클러스터로 구성되는데 예를 들면 아래 표 4와 같이 ‘꽃’과 관련된 문장들이 하나의 클러스터로 구성된다.



젤 좋아하는 꽃이 핑크튤리비 인데 ..💜💕 #빈꽃사랑  
 추운 겨울날 베란다에서 꽃을 피운  
 비 그치면 몽글몽글 피어날 겹벚꽃을 기대하며 💕  
 드디어 봄이다! 오빠랑 꽃구경 🌸  
 보라카이 에서 본 꽃인데  
 꽃보러 2시간 산넘고 물건너 와쓰옴  
 안농 벚꽃 내년에 보자🌸 주인아 귀찮다☹  
 마들렌 위로, 햇빛 비♥□

<표 4> ‘꽃’ 클러스터 문장 표

한 클러스터 내에 있는 유사한 개체명을 가진 문장들을 다수의 작업자가 주석 작업을 함으로써 다음에서 설명할 점진적 처리 프로세스에서 일관성을 확인하여 검토하는 것이 가능하다. 그 외에 전처리 도구에서는 주석 작업이 불필요한 문장들(기호 등과 같은 문장 구성)을 확인하고 제외하는 등의 기능을 제공한다.

## □ 점진적 처리 프로세스

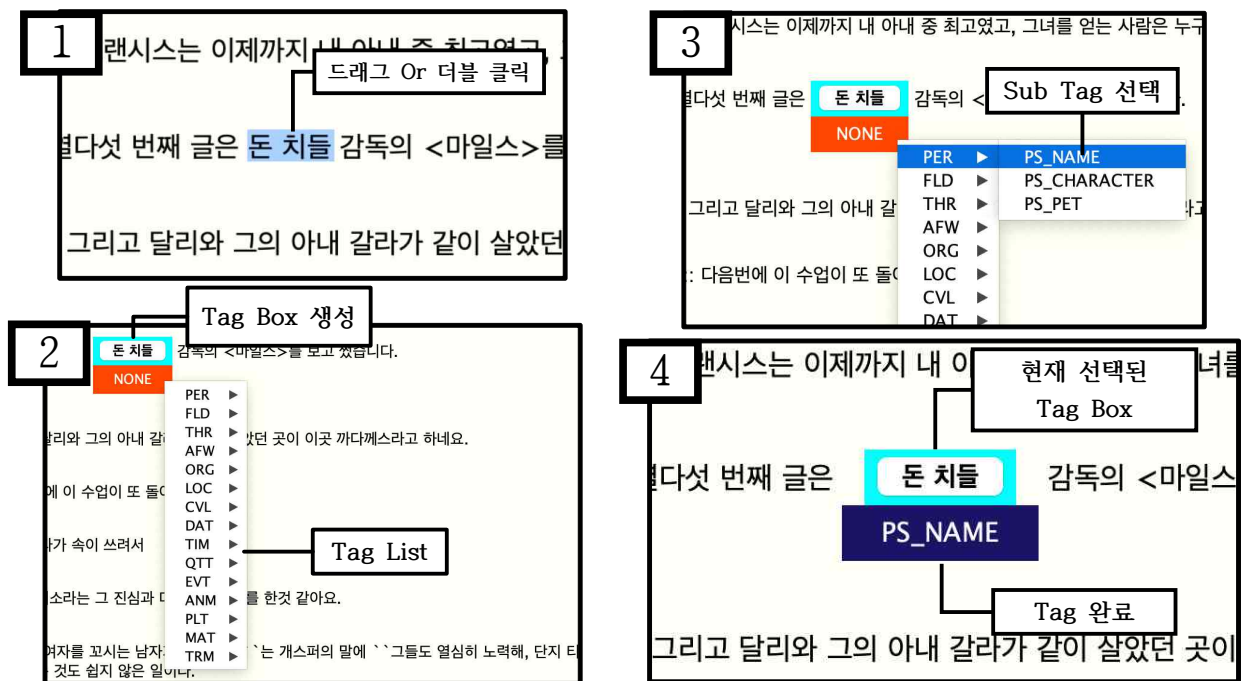
전처리 도구를 통해 얻은 클러스터를 기반으로 작업자와 시스템에 분배하여 주석 작업을 진행한다. 클러스터를 기반으로 분배를 하는 이유는 같은 클러스터 내에 표현된 개체명은 같은 분류를 가질 확률이 높기 때문이다. 위 클러스터 예와 같이 ‘꽃’과 관련된 클러스터 내에서 ‘핑크 튤립’, ‘겹벚꽃’, ‘마들렌’은 식물 개체명 분류이다. 클러스터 내 비교를 통해 다수 작업자 및 시스템에 배분된 결과 중 일관성이 떨어지는 항목 분류가 가능하다.

진행 과정을 세부적으로 설명하면 다음과 같다. 일정 기간 (약 일주일) 작업할 클러스터를 먼저 선택한다. 한 주석 작업자에게 분배되는 클러스터가 100개라고 가정한다면 나머지 주석 작업자와 시스템도 같은 100개의 클러스터를 분배받는 것이다. 작업자와 시스템은 같은 클러스터에 포함된 같지 않은 문장에 대해 주석 작업을 진행한다. 이후 주석 작업이 완료된 문장은 클러스터를 기준으로 분류하여 오류를 탐지하고 탐지된 오류를 기반으로 재수정을 진행하거나 교육을 시행한다.

이러한 프로세스를 여러 번 반복하는 과정을 점진적인 처리 프로세스로 정의하고 이 과정을 통해 말뭉치 구축을 위한 지침을 정리하고 대량의 개체명 분석 말뭉치의 품질도 높일 수 있다.

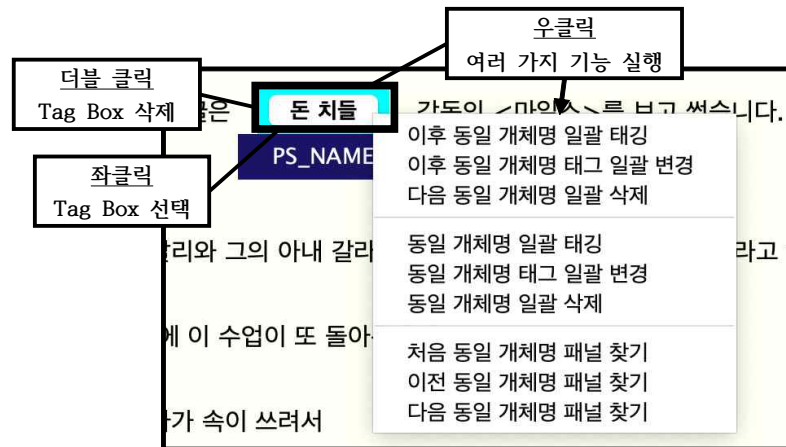
## □ 말뭉치 구축 및 수정 작업 방법 - kane 툴

말뭉치 구축 작업 시 시간 단축과 정확성을 위하여 개체명 부착 및 수정 프로그램 ‘kane’을 사용한다. ‘kane’은 말뭉치 구축 작업자가 선택한 개체명에 분류 체계에 따른 150가지 세부분류 중 적절한 하나의 태그(tag)를 부착하도록 하는 데이터리 자체 제작 프로그램이다. 자동으로 분석한 결과물 중에서 분석 모호성이 있는 문장들을 선별하여 분배하면 구축 작업자가 kane 툴을 이용하여 적절한 태그를 부착한다.

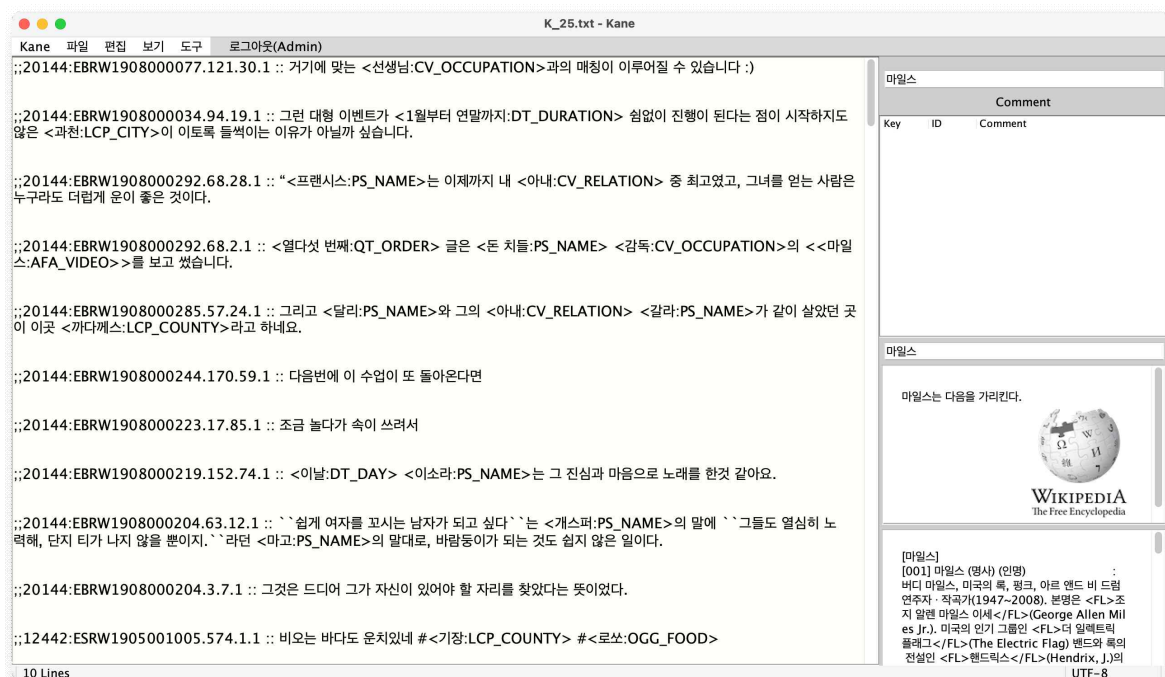


<그림 2> 구축 도구 kane 사용 방법

사용 방법은 그림 2의 순서와 같다. kane에서 작업할 문서를 연 다음, 문서에 포함된 개체명을 선택하여 개체명에 부합하는 태그를 선택하는 것이다. 개체명을 선택하는 방법은 문장에서 개체명에 해당하는 구간을 드래그하거나, 대상 어절 전체가 개체명일 때는 더블 클릭을 하면 된다. 그러면 선택된 구간에는 두 개의 박스(box)가 표시된다. 하나의 박스는 선택된 개체명 어휘를 나타내고, 나머지는 개체명 분류 태그를 보여주는데, 대분류 15가지 범주 중 하나를 선택한 후 그 범주 내의 세부분류 중 하나를 선택할 수 있다. 태그가 부착된 개체명은 문장 내에서 개체명 작업 여부를 직관적으로 확인할 수 있도록 개체명이 아닌 것과 구분하여 표시된다.



개체명 구간이 잘못 선택된 경우 더블 클릭을 하여 제거할 수 있다. 또한, 현재 작업 문서 내에 동일 개체명이 여러 개 나타날 경우 일괄적으로 태깅, 검색, 변경, 삭제할 수 있는 기능을 제공하여 작업자가 빠르고 간단하게 작업할 수 있다.



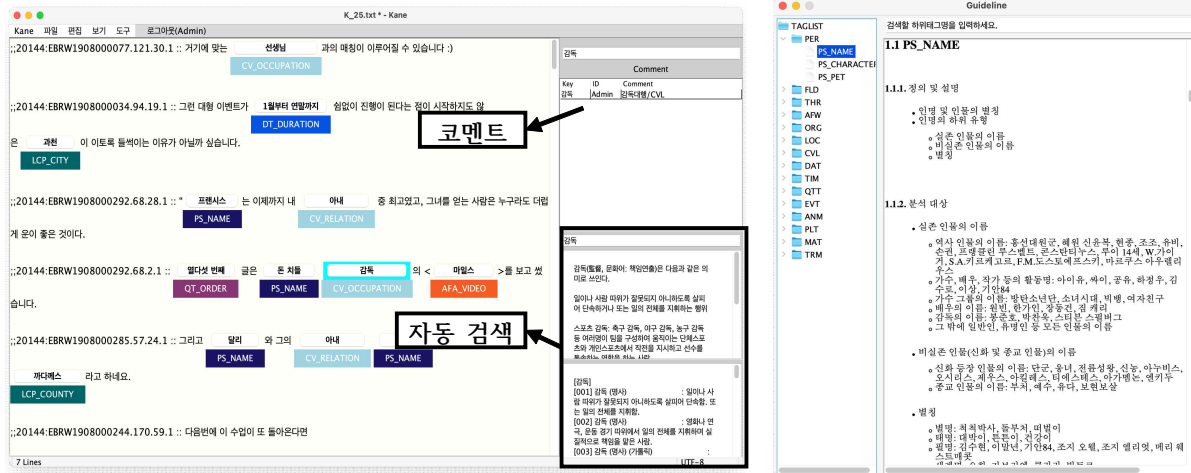
구축 작업은 한 문서를 한 명이 작업하는 것이 아니라 여러 개의 문서를 여러 명의 작업자가 하기 때문에 일관성 있는 작업 결과물이 중요하다. 이를 위하여 Kane 툴에서는 다음과 같은 기능을 제공한다.

첫 번째는 코멘트 기능이다. 작업 창에서 개체명을 선택하거나 코멘트 검색 창에 개체명을 입력하면 오른쪽 최상단에 코멘트 기능이 활성화된다. 코멘트 창에서는 선택하거나 입력한 개체명의 태깅 결과들을 검색하여 보여주기 때문에 작업자가 현재 태깅하려고 하는 개체명의 태그를 결정하는 데 도움을 받을 수 있다.

두 번째 기능은 사전 정보 자동 검색 기능이다. 개체명 작업 시 사전을 참조하여 의미, 정의를 확인하는 일이 많다. 작업자가 선택하거나 입력한 개체명의 사전 정보를 자동으로 위키백과와 우리말샘에서 검색하여 제시함으로써 작업자에게 편의를 제공한다.

마지막 기능은 가이드라인 기능이다. 개체명 단어를 선택하고 태그를 정할 때 많은 세부분류가 지침에 따라 정해지기 때문에 지침을 참고해야 하는 상황이 많이 있다. 가이드라인은 작업의 기준이 되는 개체명 분석 지침을 추가 창으로 띄워서 태깅 작업과 동시에 지침을 확인할 수 있게 해준다. 가이드라인 창에서 단어 검색이 가능하며, 가이드라인 자동 스크롤 기능이 있어 작업자가 태그를 선택하면 동시에 자동으로 해당 분류의 지침 내용으로 이동하기 때문에 작업자가 지침을 빠르게 확인할 수 있다.

'kane'을 사용함으로써 작업자들은 화면 속 작업 창에서 마우스 클릭으로 개체명 작업을 수행하므로 입력 오류를 줄일 수 있다. 또한, 위에서 설명한 다양한 편의 기능이 포함되어 있으므로 부차 시간과 오류를 줄이는 데 효과적이다. 또 작업 창에서 개체명을 선택하는 경우 오른쪽 창에 해당 개체명에 대한 코멘트와 사전 정보를, 별도 창으로 지침 정보를 실시간으로 나타내어 개체명 부차 작업의 효율을 높인다. 개인적으로 작업이 이뤄지지만, 공동 작업자들이 다른 작업자들의 결과물을 확인할 수 있어 협업이 가능했다.



<그림 5> kane의 세부 기능

본 사업에서의 ‘kane’ 툴 사용은 개체명 부착에 대한 오류 감소 및 시간을 단축하는 데 아주 중요한 역할을 했다. 본 도구를 이용함으로써 구축 작업의 정확도와 능률이 월등히 높아졌으며 사업 진행에 큰 도움이 되었다.

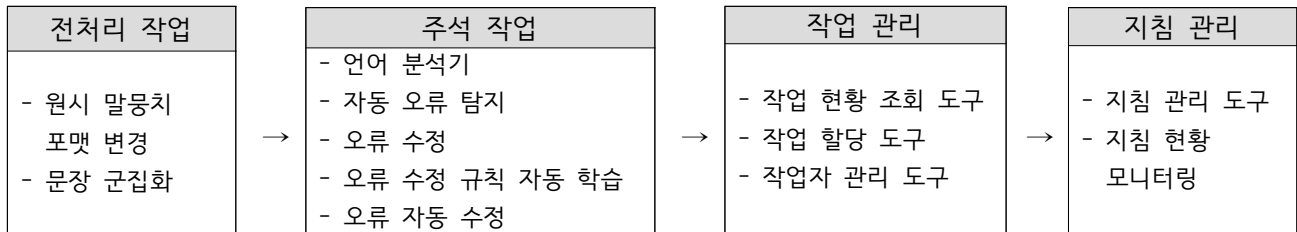
## □ 말뭉치 정책 적합 여부 검토 도구

말뭉치 구축 중에 지침 변경이 일어나는 경우가 발생한다. 이때 기존 지침에 따라 작업을 한 결과물을 변경 지침에 따라 수정하는 작업이 필요하다. 이 작업을 기존 프로세스로 진행할 경우 많은 시간이 소요되며, 구축 작업자의 혼란도가 높아져 오류 발생 가능성도 커진다.

이 문제를 해결하기 위해 변경된 지침의 반영이 필요한 부분을 검색하고, 일치되는 부분을 전문가가 검토하고 최종적으로 확정된 부분을 일괄적으로 반영을 함으로써 변경 작업에 드는 시간과 오류를 줄일 수 있다.

이 작업은 변경 지침의 적합성 검토와 적용뿐만 아니라 최종 검토 시 검토 목록을 작성하여 반영 여부를 확인하고 최종 결과물의 품질을 높인다.

### 3 말뭉치 구축 작업 과정



<표 5> 말뭉치 구축 과정

구축팀은 본 과제를 진행하며 고품질의 산출물을 위해 말뭉치를 구축함과 동시에 결과물을 점진적으로 개선하기 위한 절차를 고안하였으며 위 표와 같은 과정으로 작업을 진행하였다. 작업 과정을 설명하면 다음과 같다.

첫 번째로 이루어지는 전처리 작업에서는 원시 말뭉치의 포맷을 변환하여 문장 단위로 자동 분리하는 도구와 유사한 개체명이 있는 문장을 모아(군집화) 작업 대상을 선정하는 도구를 이용하여 작업자들에게 전달되기 전 문서를 정제하는 작업을 시행한다. 본 작업을 통해 작업자들의 편의성을 증진하여 주석 작업이 원활히 이루어질 수 있도록 준비한다. 다음으로 주석 작업에서는 먼저 언어 분석기(자동화 시스템)의 결과를 언어 문장의 작업 순서를 결정한다. 언어 분석기는 학습하는 알고리즘에 따라 분석 결과가 다르므로 다른 알고리즘으로 학습된 언어 분석기를 사용하여 얻어진 결과는 분석기 결과 간의 일치 여부를 비교할 수 있다. 일치도가 낮은 문장은 언어 분석기가 분석하기 어려운 문장으로 해석할 수 있다. 문맥에 따라 개체명 분류가 다른 경우 분석기에서 가장 혼란스러워하는 부분이다. 일치도는 문장에 대한 오류율로 정의하고, 오류율이 높은 순으로 정렬하여 작업자에게 분배하여 주석 작업자가 수정하는 작업을 거친다. 변경된 주석 문장은 언어 분석기에 학습시켜 성능을 높인다. 이후 오류율을 재계산하는 오류율 탐지와 작업자 수정 과정을 반복하여 결과물의 오류율을 감소시킨다.

작업 관리 과정에서는 통합 관리자가 개체명 및 작업자별 작업 현황을 조회하고 이에 따라 작업자에게 줄 작업 할당량을 정하여 자동 또는 수동으로 작업물을 부여한다. 또한, 작업자가 추진 일정 속도에 맞추어 주석 작업을 잘 이행하고 있는지 관리한다.

마지막으로 지침 관리 과정에서는 작업 진행 중 지침 변경 사항을 버전별로 관리하고 작업자들에게 공지하여 재교육을 진행한다.

이러한 세부 과정을 통해 작업자들의 작업 과정을 체계화하여 관리하고 말뚱치의 완성도를 높인다. 본 과정을 통해 실제 진행된 구축 작업 내용과 평가 말뚱치 구축 내용은 다음 표와 같다.

개체명 부착 작업자 작업 방식	
도구	개체명 부착 도구 Kane, 클라우드 환경 도구 활용
작업자	주석 작업자, 검토 작업자 일부
분량	전체 작업자를 기준으로 1회에 40만 어절을 배포하고, 작업 기간은 5~6일
방식	<ul style="list-style-type: none"> <li>- 같은 군집화(유사한 개체명이 포함된 문장)에 포함된 문장을 여러 작업자가 확인</li> <li>- 품질을 평가할 때 다면평가를 진행 (다면평가는 다수의 평가자가 여러 방면에서 평가하는 방법으로 3개의 언어 분석기와 같은 군집화 문장 내 포함된 개체명을 여러 작업자의 결과로 비교하는 방식을 말한다.)</li> </ul>
평가 시기	<ul style="list-style-type: none"> <li>- 한 달에 한 번 평가 말뚱치와 비교하여 수정 작업 진행</li> <li>- 누적 작업량이 40%가 넘는 시점부터 매주 자동화 시스템을 통해 도출된 오류 사항에 대한 수정 작업이 이루어짐.</li> </ul>

<표 6> 개체명 부착 작업자 작업 방식

검토/평가 작업자 작업 방식	
도구	개체명 부착 도구 Kane, 클라우드 환경 도구 활용
분량	<ul style="list-style-type: none"> <li>- 500만 어절 중 10%를 선별 평가용으로 구축</li> <li>- 주석 작업 평가를 위하여 초벌 태깅은 한 달 내에 수행</li> </ul>
방식	<ul style="list-style-type: none"> <li>- 작업자가 전문가이므로 다면 평가가 아닌 주석 도구를 이용하여 작업</li> <li>- 비정기적 작업: 변경된 지침을 반영하여 평가 말뚱치 수정</li> </ul>

<표 7> 검토/평가 작업자 작업 방식

차수	일시	작업 명	작업량 (단위:문장)	누적 작업량 (단위:문장)
1	6월 마지막 주 (20/06/29 ~ 20/07/04)	교육 1	1,500	-
2	7월 첫째 주 (20/07/06 ~ 20/07/11)	교육2	1,500	-
3	7월 둘째 주 (20/07/13 ~ 20/07/18)	작업1	60,000	60,000
4	7월 셋째 주 (20/07/20 ~ 20/07/24)	작업2	60,000	120,000
5	7월 마지막 주 (20/07/27 ~ 20/07/31)	작업3	60,000	180,000
6	8월 첫째 주 (20/08/03 ~ 20/08/07)	작업4	-	180,000
7	8월 둘째 주 (20/08/10 ~ 20/08/14)	작업5	60,000	240,000
8	8월 셋째 주 (20/08/17 ~ 20/08/21)	작업6	60,000	300,000
9	8월 마지막 주 (20/08/24 ~ 20/08/28)	작업7	55,036	355,036

10	9월 첫째 주 (20/08/31 ~ 20/09/04)	작업8	56,000	411,036
11	9월 둘째 주 (20/09/07 ~ 20/09/11)	작업9	56,000	467,036
12	9월 셋째 주 (20/09/14 ~ 20/09/18)	작업10	-	467,036
13	9월 마지막 주 (20/09/21 ~ 20/09/29)	국어원 시스템	-	467,036
14	10월 첫째 주 (20/10/05 ~ 20/10/09)	작업11	57,150	524,186
15	10월 둘째 주 (20/10/12 ~ 20/10/16)	작업12	58,000	582,186
16	10월 셋째 주 (20/10/19 ~ 20/10/23)	작업13	58,000	640,186
17	10월 마지막 주 (20/10/26 ~ 20/10/30)	작업14	58,000	698,186
18	11월 첫째 주 (20/11/02 ~ 20/11/06)	작업15	56,000	754,186
19	11월 둘째 주 (20/11/09 ~ 20/11/13)	작업16	56,000	810,186
20	11월 셋째 주 (20/11/16 ~ 20/11/20)	작업17	55,860	866,046
21	11월 마지막 주 (20/11/23 ~ 20/11/27)	작업18	49,700	915,746
22	12월 첫째 주 (20/11/30 ~ 20/12/04)	작업19	56,000	971,746
23	12월 둘째 주 (20/12/07 ~ 20/12/11)	작업20	56,000	1,027,746
24	12월 셋째 주 (20/12/14 ~ 20/12/19)	작업21	56,000	1,083,746
25	-	평가 말뭉치	103,388	1,187,134

<표 8> 말뭉치 구축 작업 내용

## 4 구축 말뭉치 품질 검증

### □ 검증 체계와 절차

구축 말뭉치의 품질을 검증하는 방법은 다면 평가 방법과 평가 말뭉치와 비교 평가, 평가 말뭉치 외 나머지 말뭉치에 대한 예측 평가가 있다.

먼저, 다면 평가는 하나의 개체명을 기준으로 여러 평가자의 결과를 비교하는 방법이다. 평가자는 앞에서 설명한 각 언어 분석기의 결과와 다수 작업자의 결과이다. 다수 작업자의 결과는 군집화를 기준으로 다면평가를 진행한다. 군집화는 유사한 개체명이 있는 문장들로 구성되어 있으므로 군집화에 포함된 문장을 다수의 작업자가 작업하여 문장에 포함된 개체명의 범주가 일치하는지를 확인하는 것이다. 다면평가 검증은 작업자가 확인이 필요한 문장을 선택하기 위해 사



용되고, 약 한 달에 한 번씩 피드백을 실행한다.

다음은 평가 말뭉치와의 비교 평가이다. 전체의 10%(약 10만 4천 문장)에 해당하는 말뭉치를 품질 검증팀과 말뭉치 구축팀이 작업하게 된다. 같은 문장에 대해 두 개의 결과가 나오므로 문장을 기준으로 개체명의 범위와 범주가 일치하는지를 계산한다. 평가 수치는 F-score를 사용한다. 아래 수식에서 평가의 기준이 되는 품질 검증팀의 결과를 answer, 성능을 비교할 말뭉치 구축팀의 결과를 (annotation) result로 표시한다. F1-measure는 F-score에서 두 가지의 평가 수치 (Precision, Recall)의 조화 평균을 구하는데 이 비율을 같게 계산하는 방식이다.

$$\text{Precision} = \frac{|\text{result} \cap \text{answer}|}{|\text{result}|}$$

$$\text{Recall} = \frac{|\text{result} \cap \text{answer}|}{|\text{answer}|}$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Precision은 result 데이터 개체명 중 기준인 answer와 일치하는 비율을 말하고 Recall은 answer의 개체명 중 result와 일치하는 비율을 뜻한다. Precision과 Recall의 조화 평균이 F1이다.

마지막 검증은 평가 말뭉치 외 예측 평가이다. 평가 말뭉치를 제외한 90%의 말뭉치는 다면평가를 통해 수행됐다. 예측 평가는 최종 말뭉치 검증 작업에서 수행되는 것으로 다면평가와 유사하다. 다면 평가가 해당 작업 회차에 대한 비교라면 예측 평가는 전체 말뭉치를 기준으로 수행한다. 수행하여 얻어진 비교 결과물은 최종 말뭉치 수정에 사용하였다.

#### □ 평가 말뭉치와의 비교 평가

위에서 설명한 두 번째 검증 방식으로 표 9는 검증 결과이다. 평가 말뭉치로 구축된 103,388문장에 대해 품질 검증팀과 말뭉치 구축팀 결과를 비교하였다.

품질 측정 결과	
Precision	94.57%
Recall	95.65%
F1	95.11%

<표 9> 품질 측정 결과

## 5 개체명 분석 말뭉치 구축 인력

### □ 사업 참여 기관



<그림 6> 사업 참여 기관

본 사업은 국가 기관인 국립국어원이 주최하는 대규모의 국가사업으로 위의 세 기관이 참여한다. 사업 기간 내내 세 기관은 끊임없는 소통과 협력을 통하여 사업을 진행해 나갔으며 각 기관이 맡은 역할을 충실히 이행하였다. 각 기관이 이행하였던 실제 과업 내용은 다음과 같다.

‘2020년 개체명 말뭉치 연구 분석’ 사업의 주최 기관인 국립국어원은 사업 이행 기관이 사업의 목적 및 목표를 뚜렷하게 인식할 수 있도록 사업의 전체적인 방향과 세부적인 내용을 명확히 설정한다. 또한, 분석 대상인 웹 말뭉치의 특성에 맞추어 새로운 범주의 태그 설정이나 세부분류 추가 및 삭제의 필요성이 발생하면 수정 사항에 대해서 최종적인 판단을 내린다. 사업의 주관으로서 용역 기관이 제출한 최종 결과물을 판단하고, 전체적인 완성도를 자세히 검토한 후 사업의 마무리를 결정한다.

데이터리는 사업 참여 기관으로서 맡게 된 ‘2020년 개체명 말뭉치 연구 분석’ 사업이 본 목적대로 잘 이행될 수 있도록 방향성을 제시하고 목표를 설정하며 성공적인 마무리를 할 수 있도록 개체명 구축 사업의 전체적인 일정과 실제

작업과 관련한 모든 단계를 직접 관리한다. 데이터리에서 보유한 자동화 시스템과 프로세스, 개체명 구축 도구, 개체명 표지 부착 인력을 이용하여 개체명 구축 작업을 진행하며, 전체 총괄 기관으로서 결과물에 대한 품질 관리와 사후에 발생하는 모든 문제에 대해 관리 책임을 갖는다.

한림대학교는 국어학적 지식을 토대로 작업한 문서의 수정과 검토를 위한 평가 말뭉치를 구축하는 동시에 지침의 문제 사항들을 잡아내고 수정하는 작업을 맡는다. 평가 말뭉치는 결과물의 내부 검증을 진행하는 과정에서 사용되며 동시에 데이터리가 보유하고 있는 도구에 반영된다.

#### □ 수행 조직 및 업무 분담

주 사업 기관은 데이터리(Data.ly)이며, 공동 사업 기관으로 한림대학교가 참여한다. 하위 부서는 사업 지원팀, 품질 검증팀, 말뭉치 구축팀, 기술 지원팀으로 이루어져 있으며 부서의 역할은 다음 설명과 같다.

사업 지원팀은 본 사업의 총괄을 맡는다. 데이터리의 대표이자 사업의 총괄을 맡은 차정원 대표가 팀장으로 있으며 사업의 시작부터 마무리까지의 진행 일정, 운영 방식, 공정 관리 등 사업 진행의 모든 절차를 관리·감독한다. 지침의 해석이나 수정에 대한 문제로 국립국어원과의 논의가 필요할 경우, 우선 본 기관의 의견을 종합하여 대표의 자격으로 국립국어원과 소통한다. 또한, 사업의 진행 방향과 일정에 차질이 없도록 사업 주최 기관인 국립국어원에 사업의 진행 척도를 주기적으로 보고하는 역할을 한다.

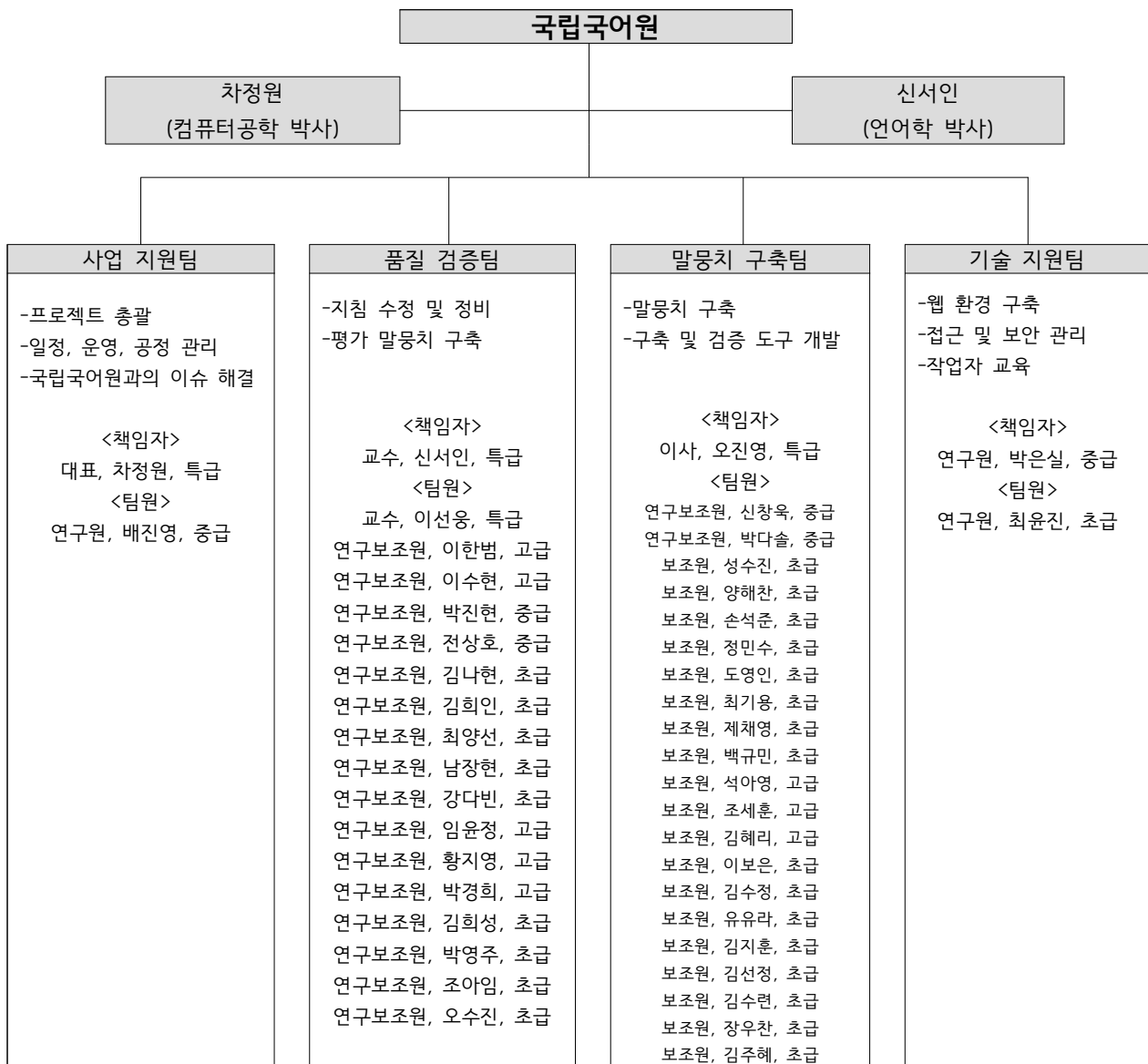
다음은 말뭉치 구축팀이다. 팀장은 데이터리의 오진영 이사이며, 말뭉치 구축에 필요한 프로세스와 도구를 개발하는 전문가들과 구축 작업을 위해 단기적으로 고용된 15명의 작업자로 이루어져 있다. 말뭉치 구축팀에서 말뭉치 구축 작업을 위해 개발한 도구에는 오류 검증 및 개체명 부착을 위한 자동화 시스템과 작업자들이 개체명 부착 작업 시 사용하는 ‘kane’이 포함된다. 말뭉치에 표지를 부착하는 작업자들은 개체명 작업 경험이 있거나 개체명 분석 지침 교육을 이수한 사람들로 이루어져 있다. 이들은 정해진 기간 안에 작업물을 완성하여 제출하여야 하며 제출한 작업물은 담당자의 검수를 받는다.

품질 검증팀은 한림대학교 신서인 교수가 팀장을 맡았다. 품질 검증팀에서는 기존의 지침에 대하여 꼼꼼히 검토한 뒤 수정 및 보완이 이루어져야 할 부분을 찾아 지침 재정비 작업을 진행한다. 지침 재정비 과정에서는 국어학적 지식을 지닌 전문가와 초기 지침을 수립한 기관인 ETRI의 자문을 받았으며, 자문 결과와 내부 검수 결과를 토대로 국립국어원과 협의하여 수정 작업을 진행하였다.

지침 수정을 거치며 지침을 완전히 숙지한 전문가들은 오류 검증을 위한 평가 말뭉치를 구축한다. 평가 말뭉치는 말뭉치 구축팀 작업물의 오류 검출 및 평가를 위해 사용되며 평가 말뭉치 구축 시 발견된 오류 및 수정 사항들을 지침에 반영하여 지침의 명확성과 구체성을 높인다.

마지막으로 기술 보증팀에서는 개체명 구축 시 사용되는 도구의 성능 확인 및 웹 환경 관리를 맡으며 외부의 접근 및 보안에 대한 관리를 맡아 사업이 철저한 보안 속에 이루어지도록 힘쓴다. 또한, 부차 작업자들의 교육 및 관리와 작업물 검수 등의 작업을 맡아 사업이 원활하게 진행되도록 한다.

앞서 설명한 수행 조직의 구조도와 사업 이행 작업자 명단은 다음과 같다.



<표 10> 2020년 개체명 말뭉치 연구 분석 사업 참여자 명단

## □ 말뭉치 구축 작업자 교육

본 사업에서 가장 중요한 부분은 개체명 구축 작업이다. 구축 작업자들은 150개의 세부분류 중 적합한 태그를 찾기 위해서는 지침을 충분히 이해하고 있어야 한다. 또한, 개체명 구축의 대상인 웹 말뭉치는 특성상 신조어 및 은어가 많고 일상에서 잘 쓰이지 않는 어휘나 표현들이 많이 등장하기 때문에 정확한 태깅 작업을 위해서는 웹 검색과 사전 정보 확인 등을 끊임없이 해야 한다. 말뭉치 구축팀에서는 이러한 점을 고려하여 작업자들이 작업 방식과 지침을 충분히 숙지할 수 있도록 교육을 실시하였다. 교육 일정과 내용은 다음과 같다.

일시	차수	진행방식	내용
20.06.29.	1차	ZOOM을 이용한 비대면 교육	<ul style="list-style-type: none"> <li>- '2020년 개체명 말뭉치 연구 분석' 사업에 대한 전반적인 안내</li> <li>- 개체명 구축 도구 'kane' 안내</li> <li>- 개체명 지침 교육 (상위 15개 분류에 대한 설명)</li> <li>- 상위 15개 분류를 이용한 태깅 작업 실습</li> </ul>
20.07.06.	2차	ZOOM을 이용한 비대면 교육	<ul style="list-style-type: none"> <li>- 개체명 지침 교육 (하위 146개 세부분류에 대한 설명)</li> <li>- 실제 작업 문서를 통한 태깅 작업 실습</li> <li>- 보안서약서, 참여확약서, 근로 계약서 작성</li> </ul>
20.08.03.	3차	대면 교육	<ul style="list-style-type: none"> <li>- 개체명의 범주 및 태깅 범위 교육</li> <li>- 지침 1.1 ver 변경 사항 안내 및 146개 세부분류 교육</li> <li>- 실제 작업 문서를 통한 태깅 작업 실습</li> <li>- 질의응답</li> </ul>
20.09.14.	4차	ZOOM을 이용한 비대면 교육	<ul style="list-style-type: none"> <li>- 지침 1.3 ver 변경 사항 안내</li> <li>- 오류가 자주 발생하는 경우들에 대한 안내</li> <li>- 이전 작업물들 수정 작업 방법 교육</li> </ul>
20.10.12.	5차	대면 교육	<ul style="list-style-type: none"> <li>- 국어원 시스템 작업 이후 성능이 50%대로 나온 작업자들만 개인 대면 교육 진행</li> <li>- 국어원 시스템 작업 결과와 개인 작업 결과 비교하여 오류 수정</li> </ul>
-	-	웹 공지	이 외 지침의 변경 및 수정 사항이 있을 시 교육 문서 작성 후 수시로 작업자들에게 공지

<표 11> 교육 진행 내역

1차 교육에서는 '2020년 개체명 말뭉치 연구 분석' 사업에 대한 소개와 구축 작업에서 쓰일 kane 도구에 대한 안내가 이루어졌다. kane을 이용하기 위한

연구원 연구참여확약서			
<b>1. 연구원 인사사항</b>			
성명	이보노	생년월일	1995-05-10
학력	한양대학교 국제학사학과 4학년 졸업		
주소	서울시 강남구 학동로 386 이치(로)2층 116-204		
연락처	(휴대전화번호) 010-3399-14201	(E-mail)	l4201@naver.com
<b>2. 참여정보</b>			
가. 참여기간 : 2020. 07. 01. ~ 2020. 12. 15.			
나. 담당업무(역할) : 설문조사 구축 (분석 결과 수록)			
다. 월 참여율 : 80%			
라. 인건비 지급액 : 월 1,190,000 원 (세금포함)			
마. 연구책임자			
성명	차 정 현		
소속 (학과 명)	데이터사이(Data.y)	직급	대표
연락처	(휴대전화번호) 056-231-3818 jcha@chungwon.ac.kr	(E-mail)	
<p>위와 같이 연구원은 연구 수행에 필요한 담당업무를 성실히 수행하며 합리하며, 재직 책임자는 이에 상응하는 인건비를 지급하며 이를 회수하여 공공으로 환원 또는 사용처가 없을 것을 확약합니다.</p> <p style="text-align: center;">2020년 7월 1일</p> <div style="display: flex; justify-content: space-between; margin-top: 20px;"> <div style="width: 45%;"> <p>연구원 : 이보노 (서명)_____</p> <p>연구책임자 : 차 정 현 (서명 또는 인)_____</p> </div> <div style="width: 45%; text-align: right;"> <p>(유자사항)</p> <p>1. 연구원에게 연구와 관련된 업무를 담당하게 하였서는 아니 됨</p> <p>2. 연구원의 담당업무(역할)을 정성충실히 수행하지 않는 경우에는 재직 책임자는 동 학칙을 위반 할 수 있음.</p> </div> </div> <p style="text-align: center; margin-top: 20px;"><b>국립국립어문장 구하</b></p>			

<그림 7> 사업 참여 관련 작업자 제출 문서

2차 교육에서는 본격적으로 146개(기존 19년 지침)의 세부분류에 대한 교육이 진행되었다. 세부 분류별 정의와 주의해야 할 사항, 예시들을 중심으로 실제 작업 문서를 이용하여 작업자들의 이해를 도왔다. 웹 말뭉치의 특성상 검색을 통하여 해당 개체명의 뜻을 파악하는 것이 중요하기 때문에 우리말샘과 사전 등을 이용한 태깅 단위 판단, 웹 검색 방법 등에 대한 교육이 진행되었다.

3차 교육은 대면으로 진행되었으며 지침 1.1 버전이 완성되었으므로 이를 중심으로 한 지침 변경 사항과 세부분류 교육이 다시 한번 진행되었다. 작업물을 검토해 보았을 때 합성어와 파생어 등의 내부 태깅 오류가 검출되었으므로 우리 말샘을 기준으로 한 내부 태깅 기준 교육을 진행하였으며 실제 문서를 통하여 작업자 전원이 함께 태깅을 진행하는 방식으로 실습을 진행하였다. 교육 진행 후에는 질의응답 시간을 가지고 작업 도구 kane에 대한 피드백과 개체명과 관

련된 질문들에 답변하였다.

지침 1.3 버전이 완성됨에 따라 4차 교육을 진행하였다. 이전의 지침들과 비교하여 새로운 세부분류 생성 및 삭제 등의 수정 사항이 많았기 때문에 이에 대한 교육과 이전 작업물에 대한 수정 작업에 대한 교육을 진행하여 지침 변경 내용이 작업물에 즉각 적용될 수 있도록 안내하였다.

마지막으로 9월 21일에 진행된 국어원 시스템 작업 결과를 토대로 태깅 오류가 많은 작업자들을 대상으로 개별 대면 교육을 시행하였다. 국어원에서 구축한 작업 결과와 작업자가 구축한 작업 결과를 비교하며 오류 수정을 진행하였으며 작업자가 주로 실수하는 부분을 분석하여 재교육하였다.

공식적인 교육 진행 이외에도 지침 변경 및 오류 사항에 대해서는 수시로 웹을 통하여 공지한 뒤 작업자들의 숙지 여부를 확인했으며 작업 중 발생하는 문제에 대해서는 적극적으로 작업자와 소통하여 해결하였다.

## □ 말뭉치 구축 작업 관리

작업자와 관리자는 작업 진행 중 소통을 위하여 승인된 이용자만이 접속할 수 있는 비공개 웹을 개설하였다. 웹에는 주요 공지사항과 kane 배포, 작업 문서 배포 안내, 작업물 제출 여부, 태깅 관련 질문 등이 게시되었다. 작업자들의 질문 중 관리자의 답변이 달린 게시글은 수집, 정리하여 웹에 공지함으로써 다른 작업자들 또한 참고하여 작업에 반영할 수 있도록 하였다.

<div>나, 너, 우리, 그, 그제,그를,저는,당신,초등학생,중학생,백범 김구,이탈리아 북부 베르나</div> <div>#60 - opened 5 days ago by 조세훈 <span>12</span></div>			
<div>태깅 질문입니다</div> <div>#51 - opened 2 days ago by 김진영 <span>12</span></div>			
<div>목/목나뭇잎 등 관련 질문 있습니다.</div> <div>#59 - opened 4 days ago by 천연주 <span>12</span></div>			
<div>자음 서열분리는</div> <div>#58 - opened 4 days ago by 조세훈 <span>12</span></div>			
<div>성명수출과 같이 수출중에는 태깅대상에 포함되나요?</div> <div>#51 - opened 4 days ago by 박상민 <span>12</span></div>			
<div>출/사/목/나뭇잎/노을/가죽 등 관련 태깅질문있습니다.</div> <div>#58 - opened 4 days ago by 천연주 <span>12</span></div>			
<div>태그 관련 질문입니다.</div> <div>#59 - opened 4 days ago by 김해리 <span>12</span></div>			
<div>어떤 태부터, 초등학생 해</div> <div>#54 - opened 4 days ago by 조세훈 <span>12</span></div>			
<div>한글문은 AFW 인가요?</div> <div>#53 - opened 4 days ago by 박상민 <span>12</span></div>			
<div>스포츠 관련 태사가 아닌 태해했</div> <div>#52 - opened 5 days ago by 장지혜 <span>12</span></div>			
<div>태그 관련 질문입니다</div> <div>#61 - opened 5 days ago by 임아림 <span>12</span></div>			
244	THR	무한 완성이 정리는 "무한성에 기초한 정리로, 타자가 앞에 있어서 마운드로 쳐대는 완성이거나 코치들이 국룰 박물관의 모든 것을 양면가하는 자 날 가능성의 거의 확실하다는 정리이다. 거의 확실하다(surely/surely)라는 말은 수학적으로 확률의 극한값이 1이라는 것을 의미한다. 위의 정리는 영어 사용자들이 특히나 세심하게 주의할 필요가 있다. 이 정리는 내용으로 적혀있다. 그리고 나오는 것 이것도 이름으로 봐야할까요? 아니면 완성이란 따로 태깅해야 할까요?	질문에 포함된 내용과 같은 문맥으로 사용한 것이라면 태깅해합니다. 태깅가 개제문으로 TML_SOCIAL_CONFERENCE로 보내요.
245	TIM	코방에서 발도 태깅하나요?	네 태깅 대상입니다.
246	CVL	수정된 지점상 cv_position 중 직장에서와 직위/직원의 위치에 따라와 사원을 잇는에 직위를 같이 따로 구분할 것입니다. "한편, 직원 경력" 같이 장르를 붙이지 않는 것은 회사 내 직책별로 어떻게 태깅하나요?	직장은 직장에 근무하는 사원을 통틀어 이르는 말로 직위/직원에 해당하지 않습니다. 그리고 마찬가지로, 임원도 위과대학에 관련된 내용이라면 태깅 대상입니다. 사원은 회사에 근무하는 사원을 통틀어 이르는 데다 직급 아래로 명칭으로 사용됩니다.
247	TMR	보조배터리, 배터리라는 TML_HVV인가요?	네 대상입니다.
248	/	한날이 한국 남자인데 '한'만 따로 태깅하나요?	지침 2.1 복합어 태깅을 확인하세요. 대상이 아닙니다.
249	/	트윈,인진 트윈터 친구, 인스타 친구인데 '트', '인'만 따로 태깅하나요?	2번과 같습니다. 그리고 역으로 정리한 일러용달을 먼저 확인해주세요.
250	CVL	앞에 메이크업, 헤어라는 단어가 없이 '아티스트'만 나뉘는 경우에도 cv_occupation인가요?	헤어와, 메이크업을 이르는 아티스트인 경우는 대상입니다.
251	CVL	가만 내에서의 en (중간) 사무총장, 운영위원, 비대위임장 등과 같은 것은 태깅하나요?	각각 단어를 직위/직책/직급으로 함께 검색해서 확인해보셔야 합니다.
252	ORG	예비군은 org_military 인가요?	네

<그림 8> 작업자 질문 사항

작업이 어느 정도 진행된 후 질의응답 부분은 웹에서 구글 시트 플랫폼으로 이동되었으며 검증팀에서 맡아 관리하였다.

작업물은 매주 월요일 오전에 배포되었으며 토요일 자정까지 제출하는 것으로

기한을 두었다. 제출 기한을 넘기거나 작업물의 품질이 좋지 않을 경우, 지각 및 경고 처리가 되며 규정에 따라 불이익이 있을 수 있음을 작업자들에게 미리 명시하여 작업에 성실히 임할 수 있도록 하였다.

2020 개체명 작업자 통계																
날짜	성명	6/29	7/6	7/13	7/20	7/27	8/3	8/10	8/17	8/24	8/31	9/7	9/14	9/21	9/28	종합
자수		교과1차	교과2차	작업1	작업2	수정작업1	대면교과	수정작업2	작업3	작업4	작업5	작업6	작업7	수정작업8	국어활용	수정작업9
A	김선정	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
B	김수현	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
C	김수정	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
D	김지훈	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
E	김혜리	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
F	도영민	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
G	백규민	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
H	석아영	○	○	○	○	○	○	○	○	○	○	○	○	○	○	1경고
I	유유라	○	○	○	○	○(지각)	○	○(지각)	○	○	○	○	○	○(경고)	○	2지각 1경고
J	이보은	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
K	장우환	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
L	채재영	○	○	○	○	○	○	○	○	○	○	○	○	○(경고)	○	1경고
M	조세훈	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
N	최기용	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
O	홍나은(입주예)	○	○	○	○	○	○(지각)	○	○	○	○	○	○	○	○	1지각
매감날짜		7/3	7/10	7/17	7/24	7/31	8/3	8/7	8/14	8/21	8/28	9/2	9/11	9/18	9/28	

<그림 9> 개체명 작업 제출 현황

## □ 품질 검증팀 업무 내용

본 사업에서 가장 중요한 부분인 개체명 구축 작업이 이루어지기 위해서는 정확한 지침을 마련하는 것이 무엇보다 중요하다. 더불어 지침을 정확히 정비하는데 있어서 가장 먼저 진행되어야 할 부분은 개체명에 대한 이해, 즉 관련 전문 지식의 습득이다. 따라서 품질 검증팀에서는 정확한 지침 마련을 위해 품질 검증팀 인력들과 주기적인 세미나를 개최하여 개체명 관련 전문 지식을 쌓는 세미나를 진행하였다. 이에 대한 세부 일정은 다음 표와 같이 이루어졌다.

순번	일시	주제	내용
1	2020.07.02.(목) 14:00~17:00	개체명	2019 국립국어원 개체명 분석 말뭉치 구축 결과보고서
2	2020.07.07.(화) 14:00~17:00	명사 의미부류	21세기 세종계획 명사 의미부류
3	2020.07.14.(화) 14:00~17:00	고유명1	박재연(2008) 고유명의 변별적 속성에 대하여 박재연(2012) 고유명 의미의 지시와 내포



4	2020.07.21.(화) 14:00~17:00	고유명2	이홍식(2016) 고유명사의 의미와 지시에 대하여 이은경(2016) 정신 공간 이론과 고유명사
5	2020.07.28.(화) 14:00~17:00	대상 부류	박동호(2000) 대상부류 개념의 응용언어학적 활용 박동호(2003) 의미부류 체계의 구축과 적용 박만규(2002) 대상부류 이론에서의 다의성
6	2020.08.18.(화) 14:00~17:00	국어 어휘의 분류	임홍빈, 한재영(1993) 국어 어휘의 분류 목록에 대한 연구 최경봉(2001) 지식기반 구축을 위한 어휘의 의미 분류 이병모(2001) 명사의 하위 분류에 대하여
7	2020.08.25.(화) 14:00~17:00	어휘의미망과 사전	고석주(2007) 어휘의미망과 사전의 뜻풀이 이동혁(2007) 의미 범주 체계의 구축과 사전에서의 활용 이동혁(2010) 사전과 어휘의미망에서의 의미 기술

<표 12> 세미나 개최 내역

이와 같이 진행된 세미나에서는 고유명 및 대상 부류에 대한 개념과 주요 논거들, 적정 술어 개념, 국어의 어휘 체계, 어휘의 의미망과 백과사전 정보와 같은 일반적 담론들을 포함하여 개체명 분석 말뭉치 구축 관련 연구보고서, 세종 전자사전 개발 등을 두루 살피며 개체명과 관련이 있는 연구들의 동향을 살폈다.

실제 작업에 있어서는 구글 스프레드시트를 활용하여 지침 마련 과정 중 팀원들과의 적극적인 소통과 즉각적인 피드백 작업이 이루어질 수 있도록 하였다. 본 사업의 개체명 대분류 체계 15종을 각각 대항목으로 나누어 시트를 구성하였고, 그 아래 ‘항목, 검토자, 검토일, 예시, 의견, 정리, 처리, 협의 결과, 협의 후 처리’ 등 9개의 세부 항목을 두어 원활하고 지속적인 지침 작성 작업이 가능토록 하였다.

개체명\_지침검토의견\_2차 ☆ 📄 📁

파일 수정 보기 삽입 서식 데이터 도구 부가기능 도움말 3분 전에 마지막으로 수정했습니다.

100% W % .0 .00 123 Arial 10 B I U A

A1	항목	검토자	검토일	예	의견	정리	처리	협의 결과	협의 후 처리
1	5.12. OGG_POLITICS	김나현	2020.10.07.	서울 시장	OGG_POLITICS 주의사항을 보면 '명사(국가, 도시, 행정구역명) + 명사' 선행하는 명사(국가 등)는 LC로 일괄 태깅한다는 문장과 함께, '서울 시장'은 <서울 LCP> <시장 CV>로 태깅한다고 되어 있으나, CV_POSITION 주의사항을 보면 <광주 시장: CV> 같이 최장 단위로 태깅한다고 되어 있어 수정이 필요합니다.	추후 지침에서 수정하도록 하겠습니다.	지침 반영 예정		지침반영1115
2									
3	5.15. OGG_OTHERS	김희인	2020.10.14	지근역비피더스 서포터즈	특정 상품의 서포터즈도 OGG_OTHERS에 태깅이 가능할까요?	OGG_OTHERS로 태깅 가능			
4	5.15. OGG_OTHERS	전상호	2020.10.17	스커러스	실존 기업이나 단체가 아닌 가상의 기업명도 태깅 대상인지 궁금합니다.	(1022)문맥을 살펴봐야 하겠으나 실존하지 않는 기업명은 태깅하지 않습니다.			
5	5.15. OGG_RELIGION	김희성	2020.10.18	대한예수교장로회	대한예수교장로회나 기독교한국침례회 같은 종교단체를 OTHERS로 태깅해야 하는지요? 둘다 종교법인단체로 알고 있고 종교사전이나 두산백과사전 등의 카테고리를 참고하면 모두 한국의 기독교 기관/단체로 분류되어 있습니다. 지침의 주의사항에 종교 및 종파의 교단은 태깅하지 않는다고 나와있는데 부모님전상서팀, 울산남매팀, 게임의 길드, 클랜까지 OGG로 태깅 가능한데 법인단체인 대형교단들을 태깅대상에서 제외한다는 것이... 다소 설득력이 부족해 보입니다. OGG_RELIGION의 종교기관, 종교단체 태깅 기준에 대한 좀더 명확한 설명이 필요한 듯 합니다.	(1022)기존의 지침에서 종교법인 단체는 태깅하지 않는 것으로 되어 있었습니다. 이부분은 후에 협의 해야 할 사항인 것 같습니다.	협의 필요	(1112 국어원회의) '대한예수교장로회, 기독교한국침례회 등과 같은 종교법인 단체 및 교단은 OGG_RELIGION으로 태깅	지침반영1115
6	5.7. OGG_MEDICINE	김희인	2020.10.21	필팩	필팩은 약을 배달하는 회사인데 이때 약국과 유사하게 보아 OGG_MEDICINE로 태깅하는지, 아니면 사기업으로 보아 OGG_ECONOMY로 태깅하는지요?	(1022)OGG_ECONOMY로 태깅합니다.			

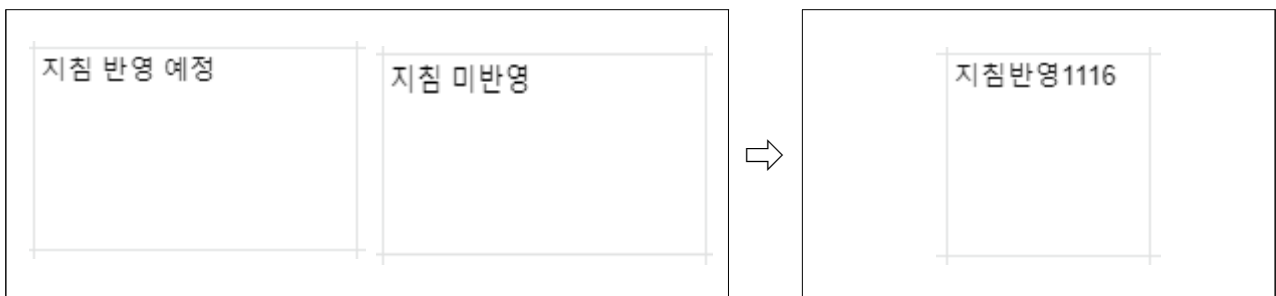
1. PERSON 2. STUDY\_FIELD 3. THEORY 4. ARTIFACTS 5. ORGANIZATION 6. LOCATIC 탐색

<그림 10> 구글 시트

1.1 PS_NAME	김희성	2020.10.18	정조대왕 이산	정조는 묘호이고 이산이 실제 이름인데 둘다 NAME으로 태깅한다면, 예와 같은 경우에는 전체를 묶어 '정조대왕 이산<PS_NAME>'과 같이 태깅하나요? 정조<PS_NAME>대왕<CV_POSITION>이산<PS_NAME> 각각 태깅하나요? (주의 사항에 보면 국대비, 폐비를 CV_POSITION으로 태깅하도록 되어 있습니다.)
-------------	-----	------------	---------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



10.20) 정조와 이산이 같은 대상은 지칭하는 것은 맞지만 하나는 묘호이고 하나는 실제 이름이라는 점에서 분리해 태깅해야 할 것 같다는 직관입니다. 따라서 '정조<PS_NAME>대왕<CV_POSITION>이산<PS_NAME>' 각각 따로 태깅하는 것이 적절해 보입니다.
---------------------------------------------------------------------------------------------------------------------------------------------------------



<그림 11> 구글 시트 활용 방법

위 그림은 실제 품질 검증팀에서 구글 스프레드시트를 활용하여 소통하는 과정을 보인 것이다. 작업자는 태깅 작업 과정 중 생기는 의문사항을 각 항목 담

당자에게 즉각적으로 질의하고, 항목 담당자는 그에 대한 의견을 적은 후 지침 반영 여부를 결정하였다. 지침에 반영한 경우 언제 반영하였는지도 기록하였다. 국어원과 협의가 필요한 사안에 대해서는 따로 표시를 하여 협의 과정을 거치도록 하였다. 품질 검증팀 내의 논의 과정 전체를 사업 주체인 국어원과 공유함으로써 문제되는 사안에 대해 함께 처리 방안을 모색할 수 있었다. 이러한 일련의 과정을 통하여 지침의 완성도를 높일 수 있도록 노력하였다.



## 제 3 장

# 개체명 분석 말뭉치 구축 지침





## 제3장 개체명 분석 말뭉치 구축 지침

### 1 지침 주요 수정 사항

본 사업에서는 2019년 국립국어원 사업 <개체명 분석 말뭉치 구축> 최종보고서에 포함되어 있는 <개체명 분석 말뭉치 구축 지침>을 정비하여 2020년 <개체명 분석 말뭉치 구축 지침>을 작성하였다. 개체명 태그 세트는 15개의 대분류와 145개의 소분류, 7개의 세분류로 구성되어 있다. 경우에 따라 소분류까지 구분되는 경우도 있고 세분류까지 구분되는 경우도 있는데 최말단 태그를 세부분류라고 부르도록 할 것이다. 최말단 태그를 모두 합하면 총 150개이다.

2019년 사업에서는 대분류 태그로 분석을 하였는데, 올해 사업에서는 세부분류 태그로 분석을 하였다. 지침을 정비하고 그 지침에 따라 분석을 한 후 실제 분석 작업 시 발견되는 문제점을 다시 지침에 반영하는 순환적인 작업을 사업 기간 내내 수행하였다. 연구원들이 공유하는 구글 시트에 지침의 각 대분류에 대한 문의 사항을 적으면 대분류 담당자가 답을 하고, 연구진 내의 논의가 필요하거나 국립국어원과의 협의가 필요하면 대면/비대면 회의 혹은 서면 회의 거쳐 지침을 수정하였다. 지침은 ver. 1.1부터 ver. 1.6까지 총 여섯 번에 걸쳐 수정하였다. 올해 사업에서 정비한 주요 수정 사항을 정리하면 다음과 같다.

2020 개체명 분류	2019 개체명 분류	비고
1. PERSON	1. PERSON	
1.1. PS_NAME	1.1. PS_NAME	
1.2. PS_CHARACTER		신설
1.3. PS_PET		신설
2. STUDY_FIELD	2. STUDY_FIELD	
2.1. FD_SCIENCE	2.1. FD_SCIENCE	
2.2. FD_SOCIAL_SCIENCE	2.2. FD_SOCIAL_SCIENCE	
2.3. FD_MEDICINE	2.3. FD_MEDICINE	
2.4. FD_ART	2.4. FD_ART	
2.5. FD_HUMANITIES	2.5. FD_PHILOSOPHY	FD_PHILOSOPHY 삭제. FD_HUMANITIES 신설.

2.6. FD_OTHERS	2.6. FD_OTHERS	
3. THEORY	3. THEORY	
3.1. TR_SCIENCE	3.1. TR_SCIENCE	
3.2. TR_SOCIAL_SCIENCE	3.2. TR_SOCIAL_SCIENCE	
3.3. TR_ART	3.3. TR_ART	
3.4. TR_HUMANITIES	3.4. TR_PHILOSOPHY	TR_PHILOSOPHY 삭제. TR_HUMANITIES 신설.
3.5. TR_MEDICINE	3.5. TR_MEDICINE	
3.6. TR_OTHERS	3.6. TR_OTHERS	
4. ARTIFACTS	4. ARTIFACTS	
4.1. AF_BUILDING	4.2. AF_BUILDING	
4.2. AF_CULTURAL_ASSET	4.1. AF_CULTURAL_ASSET	
4.3. AF_ROAD	4.4. AF_ROAD	
4.4. AF_TRANSPORT	4.6. AF_TRANSPORT	
4.5. AF_MUSICAL_INSTRUMENT	4.3. AF_MUSICAL_INSTRUMENT	
4.6. AF_WEAPON	4.5. AF_WEAPON	
4.7. AF_ART_WORKS	4.7. AF_WORKS	명칭 변경
4.7.1. AFA_DOCUMENT	4.7.1. AFW_DOCUMENT	명칭 변경
4.7.2. AFA_PERFORMANCE	4.7.2. AFW_PERFORMANCE	명칭 변경
4.7.3. AFA_VIDEO	4.7.3. AFW_VIDEO	명칭 변경
4.7.4. AFA_ART_CRAFT	4.7.4. AFW_ART_CRAFT	명칭 변경
4.7.5. AFA_MUSIC	4.7.5. AFA_MUSIC	명칭 변경
4.8. AF_WARES	4.8. AF_WARES	
4.8.1. AFW_SERVICE_PRODUCTS		신설
4.8.2. AFW_OTHER_PRODUCTS		신설
5. ORGANIZATION	5. ORGANIZATION	
5.1. OGG_ECONOMY	5.1. OGG_ECONOMY	
5.2. OGG_EDUCATION	5.2. OGG_EDUCATION	
5.3. OGG_MILITARY	5.3. OGG_MILITARY	
5.4. OGG_MEDIA	5.4. OGG_MEDIA	
5.5. OGG_SPORTS	5.5. OGG_SPORTS	
5.6. OGG_ART	5.6. OGG_ART	
5.7. OGG_MEDICINE	5.7. OGG_MEDICINE	
5.8. OGG_RELIGION	5.8. OGG_RELIGION	
5.9. OGG_SCIENCE	5.9. OGG_SCIENCE	

5.10. OGG_LIBRARY	5.10. OGG_LIBRARY	
5.11. OGG_LAW	5.11. OGG_LAW	
5.12. OGG_POLITICS	5.12. OGG_POLITICS	
5.13. OGG_FOOD	5.13. OGG_FOOD	
5.14. OGG_HOTEL	5.14. OGG_HOTEL	
5.15. OGG_OTHERS	5.15. OGG_OTHERS	
6. LOCATION	6. LOCATION	
6.1. LCP_COUNTRY	6.1. LCP_COUNTRY	
6.2. LCP_PROVINCE	6.2. LCP_PROVINCE	
6.3. LCP_COUNTY	6.3. LCP_COUNTY	
6.4. LCP_CITY	6.4. LCP_CITY	
6.5. LCP_CAPITALCITY	6.5. LCP_CAPITALCITY	
6.6. LCG_RIVER	6.6. LCG_RIVER	
6.7. LCG_OCEAN	6.7. LCG_OCEAN	
6.8. LCG_BAY	6.8. LCG_BAY	
6.9. LCG_MOUNTAIN	6.9. LCG_MOUNTAIN	
6.10. LCG_ISLAND	6.10. LCG_ISLAND	
6.11. LCG_CONTINENT	6.11. LCG_CONTINENT	
6.12. LC_SPACE	6.13. LC_SPACE	
6.13. LC_OTHERS	6.14. LC_OTHERS	
	6.12. LC_TOUR	삭제
7. CIVILIZATION	7. CIVILIZATION	
7.1. CV_CULTURE	7.1. CV_NAME	명칭 변경
7.2. CV_TRIBE	7.2. CV_TRIBE	
7.3. CV_LANGUAGE	7.8. CV_LANGUAGE	
7.4. CV_POLICY	7.5. CV_POLICY	
7.5. CV_LAW	7.18. CV_LAW	
7.6. CV_CURRENCY	7.16. CV_CURRENCY	
7.7. CV_TAX	7.6. CV_TAX	
7.8. CV_FUNDS	7.7. CV_FUNDS	
7.9. CV_ART		신설
7.10. CV_SPORTS	7.3. CV_SPORTS	
7.11. CV_SPORTS_POSITION		신설
7.12. CV_SPORTS_INST	7.4. CV_SPORTS_INST	
7.13. CV_PRIZE	7.17. CV_PRIZE	
7.14. CV_RELATION	7.14. CV_RELATION	



7.15. CV_OCCUPATION	7.15. CV_OCCUPATION	
7.16. CV_POSITION	7.13. CV_POSITION	
7.17. CV_FOOD	7.10. CV_FOOD	
7.18. CV_DRINK	7.11. CV_DRINK	
7.19. CV_FOOD_STYLE	7.19. CV_FOOD_STYLE	
7.20. CV_CLOTHING	7.12. CV_CLOTHING	
7.21. CV_BUILDING_TYPE	7.9. CV_BUILDING_TYPE	
8. DATE	8. DATE	
8.1. DT_DURATION	8.1. DT_DURATION	
8.2. DT_DAY	8.2. DT_DAY	
8.3. DT_WEEK		신설
8.4. DT_MONTH	8.3. DT_MONTH	
8.5. DT_YEAR	8.4. DT_YEAR	
8.6. DT_SEASON	8.5. DT_SEASON	
8.7. DT_GEOAGE	8.6. DT_GEOAGE	
8.8. DT_DYNASTY	8.7. DT_DYNASTY	
8.9. DT_OTHERS	8.8. DT_OTHERS	
9. TIME	9. TIME	
9.1. TI_DURATION	9.1. TI_DURATION	
9.2. TI_HOUR	9.2. TI_HOUR	
9.3. TI_MINUTE	9.3. TI_MINUTE	
9.4. TI_SECOND	9.4. TI_SECOND	
9.5. TI_OTHERS	9.5. TI_OTHERS	
10. QUANTITY	10. QUANTITY	
10.1. QT_AGE	10.1. QT_AGE	
10.2. QT_SIZE	10.2. QT_SIZE	
10.3. QT_LENGTH	10.3. QT_LENGTH	
10.4. QT_COUNT	10.4. QT_COUNT	
10.5. QT_MAN_COUNT	10.5. QT_MAN_COUNT	
10.6. QT_WEIGHT	10.6. QT_WEIGHT	
10.7. QT_PERCENTAGE	10.7. QT_PERCENTAGE	
10.8. QT_SPEED	10.8. QT_SPEED	
10.9. QT_TEMPERATURE	10.9. QT_TEMPERATURE	
10.10. QT_VOLUME	10.10. QT_VOLUME	
10.11. QT_ORDER	10.11. QT_ORDER	
10.12. QT_PRICE	10.12. QT_PRICE	

10.13. QT_PHONE	10.13. QT_PHONE	
10.14. QT_SPORTS	10.14. QT_SPORTS	
10.15. QT_CHANNEL	10.15. QT_CHANNEL	
10.16. QT_ALBUM	10.16. QT_ALBUM	
10.17. QT_ADDRESS	10.17. QT_ZIPCODE	QT_ZIPCODE 삭제. QT_ADDRESS 신설.
10.18. QT_OTHERS	10.18. QT_OTHERS	
11. EVENT	11. EVENT	
11.1. EV_ACTIVITY	11.1. EV_ACTIVITY	
11.2. EV_WAR_REVOLUTION	11.2. EV_WAR_REVOLUTION	
11.3. EV_SPORTS	11.3. EV_SPORTS	
11.4. EV_FESTIVAL	11.4. EV_FESTIVAL	
11.5. EV_OTHERS	11.5. EV_OTHERS	
12. ANIMAL	12. ANIMAL	
12.1. AM_INSECT	12.1. AM_INSECT	
12.2. AM_BIRD	12.2. AM_BIRD	
12.3. AM_FISH	12.3. AM_FISH	
12.4. AM_MAMMALIA	12.4. AM_MAMMALIA	
12.5. AM_AMPHIBIA	12.5. AM_AMPHIBIA	
12.6. AM_REPTILIA	12.6. AM_REPTILIA	
12.7. AM_TYPE	12.7. AM_TYPE	
12.8. AM_PART	12.8. AM_PART	
12.9. AM_OTHERS	12.9. AM_OTHERS	
13. PLANT	13. PLANT	
13.1. PT_FRUIT	13.1. PT_FRUIT	
13.2. PT_FLOWER	13.2. PT_FLOWER	
13.3. PT_TREE	13.3. PT_TREE	
13.4. PT_GRASS	13.4. PT_GRASS	
13.5. PT_TYPE	13.5. PT_TYPE	
13.6. PT_PART	13.6. PT_PART	
13.7. PT_OTHERS	13.7. PT_OTHERS	
14. MATERIAL	14. MATERIAL	
14.1. MT_ELEMENT	14.1. MT_ELEMENT	
14.2. MT_METAL	14.2. MT_METAL	
14.3. MT_ROCK	14.3. MT_ROCK	
14.4. MT_CHEMICAL	14.4. MT_CHEMICAL	

15. TERM	15. TERM	
15.1. TM_COLOR	15.1. TM_COLOR	
15.2. TM_DIRECTION	15.2. TM_DIRECTION	
15.3. TM_CLIMATE	15.3. TM_CLIMATE	
15.4. TM_SHAPE	15.4. TM_SHAPE	
15.5. TM_CELL_TISSUE_ORGAN	15.5. TM_CELL_TISSUE	명칭 변경
15.6. TMM_DISEASE	15.6. TMM_DISEASE	
15.7. TMM_DRUG	15.7. TMM_DRUG	
15.8. TMI_HW	15.8. TMI_HW	
15.9. TMI_SW	15.9. TMI_SW	
15.10. TMI_SITE	15.10. TMI_SITE	
15.11. TMI_EMAIL	15.11. TMI_EMAIL	
15.12. TMI_MODEL	15.12. TMI_MODEL	
15.13. TMI_SERVICE	15.13. TMI_SERVICE	
15.14. TMI_PROJECT	15.14. TMI_PROJECT	
15.15. TMIG_GENRE	15.15. TMIG_GENRE	
15.16. TM_SPORTS	15.16. TM_SPORTS	

<표 13> 지침 주요 수정 사항

## 2 개체명 분석의 대원칙 정비

개체명 분석은 태깅 대상 파악, 태깅 단위 판단, 태깅 범주 결정의 세 단계로 나누어진다. 기존 지침에서는 이에 대한 논의가 지침 뒷부분에 있었는데 이 내용들은 총론에 해당하므로 지침 앞부분으로 가져와 전체에 적용되는 사항들임을 분명히 하였다.

첫째, ‘태깅 대상’은 ‘무엇을 분석할 것인가’의 문제이다. 개체명 태깅의 대상은 기본적으로 고유명이지만, 자연종명이나 상품명 등 고유명인지 보통명사인지 구분이 쉽지 않은 개체들도 있고, 보통명사 중에도 실용적인 목적으로 개체명 태깅이 필요한 것들도 있다. 따라서 ‘태깅 대상’ 부분에서는 개체명 태깅 대상으로 삼는 명사 부류를 고유명과 보통명사로 나누어 그 유형을 나열하였다.

둘째, ‘태깅 단위’는 ‘어디서부터 어디까지를 분석할 것인가’의 문제이다. 태깅 단위 결정 원리를 고유명과 보통명사로 나누어 논의하였다. 태깅 단위 결

정의 대원칙은 ‘최장 단위 태깅’이다. 즉, 하나의 개체를 가리키는 언어 표현은 하나의 개체명으로 보고 태깅한다는 것이다. 그러나 실용적인 목적을 위하여 내부 구성요소의 정보를 참조할 필요가 있는 경우에는 예외적으로 ‘최소 단위 태깅’을 적용하도록 하였다. ‘최장 단위 태깅’과 ‘최소 단위 태깅’이라는 두 원칙 사이에 혼동이 없도록 예외적인 경우를 분명히 명시하였다.

셋째, ‘태깅 범주’는 ‘무엇으로 분석할 것인가’의 문제이다. 단어나 구 중에는 둘 이상의 개체명 범주에 해당하는 것들이 있다. 태깅 범주 결정의 대원칙은 문맥에 따라 결정한다는 것이다. 하나의 단어가 둘 이상의 범주에 속할 때에는 적정술어를 활용하여 결정한다. 예를 들어, ‘시금치’가 ‘심다, 키우다, 뽑다’나 ‘자라다, 시들다’ 등의 서술어와 함께 사용되었으면 PT\_GRASS로 판정하고, ‘데치다, 볶다, 무치다, 요리하다, 먹다’ 등의 서술어와 함께 사용되었으면 CV\_FOOD로 판정하는 식이다. 한편 문맥을 통해 해당 개체명의 의미 범주를 판별할 수 없는 경우에는 우선순위를 제한적으로 적용하여 태깅하도록 하였다.

그 밖에 특수 항목 태깅으로 괄호 안의 원어 정보나 부가 정보, 원어절 사이에 기호가 삽입되어 있는 경우, 날짜/시간/수량 명사 + 보조사/접사/단위성 의존 명사, 영어나 한자 등의 외국어, 줄임말, 개체명의 일부로 해당 개체를 가리키는 경우, 비유적 표현, 개체명과 조사/어미가 하나의 음절로 실현된 경우, 오자나 오타 등에 대해 태깅 지침을 자세히 제시하였다.

### 3 세부분류별 분석 지침 정비

#### □ 자문 내용

지침의 완성도를 높이기 위하여 국어학적 지식을 지닌 전문가들과 본 개체명 사업의 기준 지침이 되었던 ‘세부분류 개체명 가이드라인 2018’을 작성한 ETRI 전문가의 자문을 받았다. 자문 내용은 다음과 같다.

#### 1. 1차 자문

- 자문위원: 배경만 박사 (ETRI)
- 자문 내용
  - 개체명 분석 대상(일반명사 태깅 기준)  
: 위키피디아 상식 분야에서 키워드가 될 만한 것들을 개체명 분석 대상으

로 선정하여 일반명사도 포함한다.

- 개체명 분석 체계  
: 기존 개체명 분석 체계는 활용할 수 있는 정도의 표준이지, 반드시 따라야 하는 표준이 아니며 개체명 분류 체계는 활용 목적에 따라 달라질 수 있다. (자체적으로 확장, 수정 가능)
- 외연이 불명확한 개체명의 태깅 방안  
: 2~3개의 태그가 가능한 경우 문맥을 기반으로 판단하는 것이 가장 기본이며, 문맥으로 판단이 모호한 경우는 일관성을 위해 우선순위를 정하여 태깅한다.
- AF\_WARES 세분화 가능성  
: 태그 세트의 체계가 어지럽다고 판단된다면 수정할 수 있지만, 개체명 태깅은 일반적으로 공학 계열에서 사용하기 때문에 세분화하지 않아도 활용에는 문제가 없다.
- 7.13. CV\_POSITION, 7.14. CV\_RELATION 7.15. CV\_OCCUPATION의 태깅 대상  
: 일반명사가 개체명이 되는 경우 기준을 명확히 하는 것이 중요하다. 기준을 명확히 하여 개체명 대상을 줄이는 과정이 필요하다.

## 2. 2차 자문

- 자문위원: 정경재 교수 (창원대학교)
- 자문 내용
  - 축소 지칭  
: ‘1.1. 축소 지칭’ 과 ‘1.3. 절단되어 일부만 쓰인 개체명’ 은 상충하므로 일관성 있게 다듬을 필요가 있다.
  - 괄호 안의 원어 정보  
: ‘괄호 안의 원어 정보가 괄호 밖의 내용과 같은 경우 태깅하지 않는다’ 라는 지침에 대해, 괄호 안의 원어 정보 또한 개체명에 해당하는 것이므로 태깅하는 것이 말뭉치 취지에 어긋나지 않고 이후에 이용하기에도 쉬울 것으로 보인다.
  - 최소 단위 태깅 원칙과 최장 단위 태깅 원칙  
: ‘최장 단위 태깅 원칙’ 을 우선으로 적용한다면, 명사구 태깅 사례만 따로 뽑아 검토할 수 있다는 이점이 있다. ‘최소 단위 태깅 원칙’ 으로 태깅 하면, 명사구로 묶어서 처리할 사례를 다시 추려내기 어렵다.
  - 유일 태깅 원칙  
: ‘문맥 우선 적용 원칙’ 에 동의하나, 어떠한 건축물(AF)은 어딘가에 위치

해 있기 때문에 특정한 장소(LC)의 의미를 가질 수밖에 없으므로 ‘우선 순위 적용 유형’으로 처리하는 방안을 검토할 필요가 있다.

- CV 범주

: CV\_POSITION의 정의에서 ‘일시적으로 맡는 역할’이라는 것은 그 영역을 규정하기 어려우므로, 이 부분은 전체를 태깅 대상에서 제외하는 것을 검토할 필요가 있다. ‘인간관계’를 의미하는 CV\_RELATION 역시 검토 대상이다.

### 3. 3차 자문

- 자문위원: 박동호 교수(경희대학교)

- 자문 내용

- 세종 명사 의미부류 작업에서 대분류-중분류-소분류를 나누는 원칙

: 세종 명사 의미부류에서는 대분류를 <구체물>, <집단>, <장소>, <추상적 대상>, <사태>의 다섯 가지로 나누었으며, 대분류는 대부분 적정 술어를 이용하여 구분하였다. 적정 술어 이외에도 다양한 형태론적 정보(예: 접미사 -인, -자 → <인물>), 통사론적 정보, 명사-명사 공기 관계 등의 정보도 활용하였다.

- 적정 술어 활용

: 세종 명사 의미부류에서는 어휘의 다의성을 인정하고 각각 다른 의미로 쓰였을 때 서로 다른 부류에 소속되게 하였다. ‘부추’의 경우 ‘기르다, 시들다’와 같은 동사와 공기하면 <식물>에, ‘재배하다’와 같은 동사와 공기하면 <채소>에 소속시킨다. 적정 술어에 따라 구분하여 서로 다른 부류에 소속시키는 처리가 언어 현실을 충실히 반영하는 것이라 생각한다.

- 세종전자사전의 명사 의미부류 활용 분야

: 대상 부류 이론에 기반한 명사 의미부류는 동사나 형용사 등 술어의 논항 선택 제약을 기술하는 데에 활용된다. 또한, 명사의 다의어를 기술할 때 해당 하위 의미 항목이 어떤 명사 의미부류에 속하는지 표시하면 그 차이를 분명하게 드러낼 수 있다.

- ‘장소’ 관련 명사 태깅 시 문제와 해결 방안

: 하나의 어휘가 건물, 기관, 장소 등의 의미를 가진다면 문맥에 따라 구분하여 각 부류에 소속시키는 것이 원칙이다. 그러나 적정 술어가 다의성을 가지는 경우도 있으므로, 적정 술어의 의미에 유의하면서 활용해야 한다.

#### 4. 4차 자문

- 자문위원: 정경재 교수 (창원대학교)

- 자문 내용

- 태깅 대상

: 유형을 나타내는 일반명사가 태깅 대상인 세분류에 누락된 것이 일부 있다. EV에서도 일반명사가 태깅되는 경우가 있는데 전반적으로 다시 검토해서 정리할 필요가 있다.

- 태깅 단위

: ‘2.2. 구의 태깅’ 부분에서 “구 단위가 사전에 등재되어 있지 않은 경우, 내부 요소를 각각 따로 태깅한다.” 라고 되어 있는데, 이는 ‘최장 단위 태깅(원칙)’ 과 충돌하는 것으로 보여 표현을 다듬을 필요가 있을 듯하다.

- 태깅 범주

: ‘2.2. 우선순위 적용(예외)’ 에서 문맥으로 범주를 판별할 수 없는 예들을 표로 정리하여 보여주고 있는데, 누락된 유형이 많이 있다. 여러 예시를 정리하여 추가할 필요가 있다.

- PERSON

: PS\_CHARACTER에 “주로 만화영화나 게임 등에 등장하는 독특한 인물” 이 포함된다고 되어 있는데, 뽀로로, 루피처럼 사람이 아닌 존재는 문제가 없으나 ‘짱구, 심슨’ 처럼 사람으로 여겨지는 존재는 PS\_NAME과 경계가 불분명하다.

- LOCATION

: ‘6.13.3’ 에 ‘울산바위, 갯바위’ 가 LC\_OTHERS의 예로 제시되어 있는데, 바위 이름을 장소로 볼 수 있을지 의문이다 ‘샘터, 약수터, 계곡, 동굴’ 등이 특정한 공간을 의미하는 것에 비해 바위는 사물이기 때문에 대상에서 제외하여야 한다.

- EVENT

: EV의 경우 특정한 시기나 연도를 상징할 수 있는 사건, 사고, 사회적 현상에 한정해서 태깅하는 것이 원칙이나 ‘주민투표, 시험, 정상회담’ 의 경우 일반명사도 태깅하는 것으로 되어 있다. ETRI 초기 지침에서부터 EV의 하위 유형에 따라 처리 방식이 다르고 산발적인 예외들이 포함돼 있었다. 지난 사업에서 이를 정리하고자 했으나 정리하지 못했고 기존 지침을 수용해서 다음과 같이 정리하는 것까지밖에 하지 못했다. 만약 이번 사업에서 지침을 전면적으로 개정할 수 있는 상황이라면 이 부분에서 큰 의미 없는 예외를 정리해 보는 것을 권유한다.

• TERM

: TM\_CELL\_TISSUE는 본래 “눈으로 볼 수 없는, 외부에서 보이지 않는 부분”을 태깅하는 것이었으며 15.5.4에 제시한 ‘태깅 범위’에서도 그렇게 명시하고 있다. 그러나 현 지침에서 예시로 ‘머리털, 털, 눈썹, 이’ 등 외부로 드러나 눈으로 볼 수 있는 것들 다수가 포함되어 있다. 정의를 좀 더 명확하게 밝혀서 ANIMAL\_PART와 명확히 구분되는 기준을 보여야 한다.

□ 회의 내용

지침 수정은 기관 간 협의가 필요하므로 다음과 같은 회의가 이루어졌다.

차수	참여 기관	회의 내용
1차	국립국어원 데이터리 한림대학교	1. PERSON - SNS 아이디 및 애완동물 이름의 태깅 범주 2. STUDY_FIELD - ‘예술 분류’로 제시된 FLD_ART 항목들의 적절성 판단 - FD_OTHERS 예시들의 적절성 판단 - 세부분류 FD_HUMANITIES 생성 찬반 3. ARTIFACTS - AFW_WARES의 분류 기준의 모호성 4. ORGANIZATION - ‘○○지점, ○○공장, ○○캠퍼스’의 경우 태깅 단위 통일 - OGG_POLITICS의 태깅 기준 - OGG_HOTEL의 태깅 범위 5. LOCATION - LC_TOUR와 LC_OTHERS의 구분 기준 - ‘사거리, 로터리, 삼거리’와 ‘산책로, 등산로’의 태깅 분류 - ‘미아사거리, 홍대’ 등 위치 일대를 일컫는 경우의 태깅 분류 6. EVENT - EV_FESTIVAL에 행사, 콘서트 등도 포함될 수 있도록 설명에 추가
2차	국립국어원 데이터리 한림대학교	1. CIVILIZATION - CV_CLOTHING의 태깅 범주 - CV_POSITION의 태깅 범주 2. DATE - DT_DURATION과 DT_OTHERS의 판단 기준 : ‘~가량’, ‘D+N일’ 등 - 시간이나 날짜에 구체성을 더하는 표현(이번~, 지난~ 등)도 태깅 단위에 포함. - DATE와 QUANTITY의 정확한 기준 3. TIME - TI_DURATION의 태깅 범주



		<ul style="list-style-type: none"> <li>- 시간에 구체성을 더하는 보조사(~만, ~쯤, ~씩 등)도 태깅 단위에 포함</li> </ul> <p>4. QUANTITY</p> <ul style="list-style-type: none"> <li>- 구체성을 더하는 보조사(~만, ~쯤, ~씩 등)도 태깅 단위에 포함</li> <li>- QT_SIZE와 QT_LENGTH의 중복 범위 해결</li> <li>- '수 관형사+개체명 또는 일반명사'의 경우 태깅 방법</li> </ul> <p>5. 태깅 단위</p> <ul style="list-style-type: none"> <li>- '우리말샘'을 기준으로 하여 판단하기로 결정</li> <li>- 'ˆ' 표시의 경우 전문어이기 때문에 한 단어로 태깅</li> <li>- '-' 표시의 경우 한 단어로 보고 내부 태깅 하지 않음.</li> </ul>
3차	국립국어원 데이터리 한림대학교	<p>1. PERSON</p> <ul style="list-style-type: none"> <li>- 'PS_NAME'의 주의 사항에 실존하는 특정 대상을 명확하게 지칭하는 경우만 태깅함을 명시</li> </ul> <p>2. ORGANIZATION/LOCATION/ARTIFACTS</p> <ul style="list-style-type: none"> <li>- 문맥과 지침에서 제시한 적정 술어만으로는 구분이 어려움.</li> <li>- 여러 예제를 추가하여 범위 제시</li> <li>- LOCATION의 경우 LC_TOUR 삭제 고려</li> </ul> <p>3. CIVILIZATION</p> <ul style="list-style-type: none"> <li>- CV_POSITION에 고대국가 지위와 스포츠 포지션 포함 : 하위범주인 CV_SPORTS_POSITION 추가하는 방안</li> </ul> <p>4. DATE/QUANTITY</p> <ul style="list-style-type: none"> <li>- '~일째'의 경우 어느 범주에 속하는가? : 날짜를 의미하는 DATE 카테고리에 포함 결정</li> </ul> <p>5. ARTIFACTS/TERM</p> <ul style="list-style-type: none"> <li>- 화장품 색상의 명칭 태깅 방법 : 상품명과 색상 명칭을 통합하여 AFW_OTHER_PRODUCTS로 태깅</li> </ul>

<표 14> 지침 수정 관련 회의 내용

## □ 세부분류별 수정 사항

위와 같은 과정을 통하여 최종적으로 각 세부분류별 분석 지침을 전면적으로 정비하였다. 모든 항목에 대해 정의 및 설명, 분석 대상, 예문, 주의 사항의 순서로 기술하였다.

### 1. PERSON

- 작품에 나오는 등장인물을 태깅하기 위해 세부분류 PS\_CHARACTER를 신설하였다.
- 반려동물의 이름을 태깅하기 위해 PS\_PET을 신설하였다. 반려동물은 동물이지만 이름을 부여하고 가족의 일원으로 취급한다는 점에서 인간과 유사한 개체로 볼 수 있다고 판단하였다.

## 2. STUDY\_FIELD

- 학문 분야 중 인문학에 해당하는 세부분류가 없어서 FD\_HUMANITIES를 신설하였다. 세부분류 FD\_PHILOSOPHY는 삭제하고 철학도 FD\_HUMANITIES으로 태깅하기로 하였다.

## 3. THEORY

- 이론 중 인문학 이론에 해당하는 세부분류가 없어서 TR\_HUMANITIES를 신설하였다. 세부분류 TR\_PHILOSOPHY는 삭제하고 철학 이론은 TR\_HUMANITIES로 태깅하기로 하였다.

## 4. ARTIFACTS

- 이전 지침에서는 AF\_CULTURAL\_ASSET의 태깅 대상으로 무형문화재, 천연기념물 등 인공물에 해당하지 않는 것들이 들어 있었는데, AF\_CULTURAL\_ASSET의 태깅 대상을 건축물, 고분 등의 문화재로 한정하였다.
- 기존 지침의 AF\_WORKS는 AF\_ART\_WORKS로 명칭을 변경하였다. AFW\_DOCUMENTS, AFW\_PERFORMANCE, AFW\_VIDEO, AFW\_ART\_CRAFT, AFA\_MUSIC은 각각 AFA\_DOCUMENTS, AFA\_PERFORMANCE, AFA\_VIDEO, AFA\_ART\_CRAFT, AFA\_MUSIC으로 명칭을 변경하였다. 이는 AF\_WARES의 하위 분류와 구분하기 위함이다.
- 공산품의 상품명과 서비스업의 상품명을 구분하기 위하여 AF\_WARES의 하위분류로 AFW\_SERVICE\_PRODUCTS, AFW\_OTHER\_PRODUCTS를 신설하였다.
- 태깅의 혼란을 방지하기 위해 AF\_WARES로 태깅하던 식음료 가공품을 모두 CV\_FOOD 혹은 CV\_DRINK로 태깅하기로 하였다.

## 5. ORGANIZATION

- OGG\_RELIGION의 태깅 대상에 종교법인단체와 교단을 포함시켰다.

## 6. LOCATION

- 세부분류 LC\_TOUR를 삭제하였다. 어떤 지역을 사람들이 많이 찾으면 관광지가 된다. 따라서 어떤 관광지이든 본래적인 범주가 있고 부가적으로 관광지라는 속성을 가지게 되는 것이다. 따라서 관광지의 태깅 대상은 늘 여

러 가지 태그로 분석될 가능성이 있었다. 또한 관광지는 총망라하기도 어렵고 태깅 작업자마다 판단이 다를 수도 있다. 그래서 기존에 LC\_TOUR로 태깅하던 대상들은 AF\_BUILDING, AF\_CULTURE\_ASSET, LC\_OTHERS 등 본래적인 범주에 따라 태깅하기로 하였다.

## 7. CIVILIZATION

- 문명이나 문화를 가리키던 CV\_NAME을 CV\_CULTURE로 명칭을 변경하였다.
- 예술의 하위분류를 태깅하기 위해 세부분류 CV\_ART를 신설하였다.
- CV\_RELATION의 태깅 대상을 가족, 친족 명사로 한정하였다.
- CV\_POSITION의 태깅 대상을 직위, 직책, 직급으로 한정하고, 스포츠 포지션을 따로 태깅하기 위해 CV\_SPORTS\_POSITION을 신설하였다.
- 기존에 AF\_WARES로 태깅하던 식음료 가공품을 CV\_FOOD나 CV\_DRINK로 태깅하기로 하였다.

## 8. DATE

- 기존의 지침에 일, 월, 연 등을 태깅하는 세부분류는 있었으나 주 단위를 태깅할 세부분류가 없어서 DT\_WEEK를 추가하였다.

## 9. TIME

- 아침, 점심, 저녁, 낮, 밤, 새벽 등의 보통명사를 TI\_DURATION의 태깅 대상에 포함시켰다.

## 10. QUANTITY

- 기존의 지침에는 QT\_ZIPCODE가 있어 우편번호를 태깅하게 되어 있었으나 주소와 관련된 그 밖의 숫자들, 즉 주소의 상세 번지나 사서함 번호 등을 태깅할 범주가 없어서 QT\_OTHERS로 태깅했었다. 이에 QT\_ADDRESS를 신설하여 사서함 번호, 우편번호, 번지, 동, 호수 등 주소에 관련된 모든 숫자를 QT\_ADDRESS로 태깅하기로 하였다.

## 11. EVENT

- EVENT는 일시와 장소가 분명한 사건, 사고, 행사 등에 한정하여 태깅하기로 하였다. 4차 산업 혁명, 르네상스 등과 같이 시기나 기간을 특정할 수

없는 사조나 사회 변화를 의미하는 명칭은 CV\_CULTURE로 태깅하기로 하였다.

- EV\_FESTIVAL의 태깅 대상에 콘서트, 행사 등을 추가하였다.
- EV\_OTHERS는 문맥상에서 일시, 장소가 명확한 특정 행사, 사건으로 볼 수 있는 항목인 경우에만 태깅하는 것으로 한정하였다.

## 12. ANIMAL

- 동물은 ANIMAL로 태깅하되 문맥에서 음식의 재료로 쓰이면 CV\_FOOD로 태깅하기로 하였다.
- 신체 부위 중 살, 피부, 털, 머리카락, 이, 이빨, 혀 등 눈으로 확인 가능한 신체 부위는 AM\_PART로 태깅하기로 하였다.

## 13. PLANT

- 식물은 PLANT로 태깅하되 문맥에서 음식의 재료로 쓰이면 CV\_FOOD로 태깅하기로 하였다.
- 식물명은 식물의 일부분을 가리키기도 하고 식물 전체를 가리키기도 하는 등 다의성을 가지는데, 적정술어를 이용해서 하위의미를 구분하여 태깅하기로 하였다. 즉, ‘산수유가 노랗게 피었다.’라는 문장의 ‘산수유’는 PT\_FLOWER로 태깅하고, ‘빨갭게 익은 산수유를 따다.’라는 문장의 ‘산수유’는 PT\_FRUIT로 태깅하고, ‘마당에 산수유를 심었다.’라는 문장의 ‘산수유’는 PT\_TREE로 태깅하는 것이다.
- <우리말샘>에 ‘꺾-꺾질’이나 ‘감꺾^꺾질’과 같이 ‘-’나 ‘^’를 포함하여 등재되어 있는 복합어는 내부 요소인 과일명을 태깅하지 않지만, 치자 추출물의 경우 <우리말샘>에 구로 등재되어 있으므로 이 구 안에서 열매를 지칭하는 ‘치자’를 ‘PT\_FRUIT’로 태깅하는 것으로 하였다.

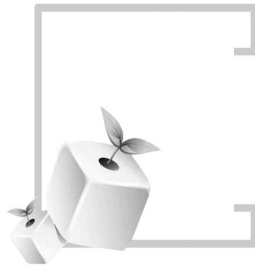
## 14. MATERIAL

- MT\_CHEMICAL의 태깅 대상을 ‘고체 화학물질, 액체 화학물질, 기체 화학물질, 영양소 및 호르몬’과 같이 구체적으로 명시하였다.

## 15. TERM

- TM\_CELL\_TISSUE로 태깅되는 대상은 세포, 기관, 조직인데, 이를 분명히 하기 위해 세부분류 명칭을 TM\_CELL\_TISSUE\_ORGAN으로 변경하였다.

- 기존에 TM\_CELL\_TISSUE\_ORGAN로 태깅하던 대상들 중 눈으로 확인 가능한 신체 부위는 AM\_PART로 태깅하기로 하였다.



## 제 4 장

# 사업 수행 결과





## 제4장 사업 수행 결과

### 1 사업 요약

국립국어원이 주관하고 기업 데이터리(Data.ly)와 공동참여기관 한림대학교가 함께 참여하여 진행한 ‘2020년 개체명 말뭉치 연구 분석’ 사업은 2020년 6월 1일부터 시작하여 2020년 12월 31일까지 약 7개월간 진행되었다. 사업의 주 내용은 웹 원시 말뭉치 500만 어절에 대한 개체명 분석과 말뭉치 부착 작업을 오류율 5% 이내로 완료하는 것과 작년 개체명 사업을 통해 완성된 ‘19년 국립국어원 개체명 분석 말뭉치 지침’ 내용을 수정 및 보완하는 것이었으며, 완료된 작업물은 민간에 배포되어 관련 연구 및 사업에 이바지하는 것을 목표로 삼았다.

사업 수행 기관인 데이터리와 한림대학교는 각자 지닌 전문적인 지식에 따라 담당 역할을 나누었고 크게 ‘말뭉치 구축팀’과 ‘검질 보증팀’으로 분류하였다. 구축팀에서는 구축 프로그램 ‘kane’을 이용한 개체명 부착 작업과 자동화 시스템을 이용한 오류 검출 및 수정 작업과 인력 관리를 맡았으며 검증팀에서는 국어학적 지식을 지닌 전문가가 지침에 대해 검토하고 문제점을 보완하는 작업과 완료된 작업물들의 검증을 위한 평가 말뭉치를 구축하는 작업을 시행하였다.

말뭉치 구축 작업 과정에서는 자체 개발한 개체명 구축 프로세스와 구축 도구 ‘kane’을 이용하여 신속하고 정확한 구축 작업을 진행할 수 있었다. 또한, 개체명 작업에 특화된 자동화 시스템을 이용하여 말뭉치 정책 적합 여부를 판단하고 평가 말뭉치와의 비교를 통해 오류를 수정하는 작업을 진행하여 최종 말뭉치의 정확도를 높였다. 컴퓨터 기술과 인력 작업을 병행하여 작업을 진행함으로써 빠르고 정확하게 구축 작업을 진행할 수 있었다.

지침 수정 및 보완의 과정에서는 ‘개체명’의 정의를 명확히 하고 태깅의 단위와 범위를 상세화하였다. 기존 지침을 면밀히 검토하면서 웹 원시 말뭉치의 특성이 반영될 수 있도록 분류 체계를 정비하고 지침의 세부 내용을 보완하였다. 이러한 과정을 거쳐 최종적으로 상위 분류 15개, 세부분류 150개를 확정하였다.

‘2020 개체명 연구 분석’ 과제는 미래 전망적인 인공지능 관련 산업과 4차 산업혁명의 발전에 기반이 될 수 있는 대량의 개체명 분석 웹 말뭉치를 구축하고 이를 위한 지침을 재정비하는 데 의의가 있다.



인터넷이 발달하면서 웹은 우리의 중요한 의사소통 환경이 되었다. 따라서 웹 환경에서 사용되는 언어 자료를 모아 구축한 말뭉치는 언어처리 산업 분야뿐만 아니라 국어 및 국어문화 연구와 국어정책 수립의 기초 자료로 매우 유용하게 사용될 수 있다. 또한, 이러한 웹 말뭉치에는 생활에 밀접한 다양한 개체명이 등장하므로 웹 말뭉치를 대상으로 개체명 정보를 부착한 개체명 분석 말뭉치는 다양한 언어 인공지능 산업에 유용하게 활용될 것이다.

본 사업에서는 개체명 분류 기준을 세분화, 구체화하고 모호한 부분을 삭제하는 등 지침을 전체적으로 재정비하였다. 말뭉치 구축 지침은 추후 관련 사업에서도 기준이 되어야 하는 중요한 자료이다. 사업을 통해 만들어진 완성도가 높은 지침은 향후 관련 연구 진행에서도 본보기로 사용되어 관련 사업이나 연구가 성공적으로 마무리될 수 있도록 도움을 줄 것이다.

본 사업을 통해 생성된 최종 작업물은 민간에 공유되며 현재 주목받고 있는 기술인 언어 인공지능 산업과 빅데이터 산업 등 관련 사업에 중요한 토대로 쓰이게 된다. 언어 인공지능 산업과 빅데이터 산업의 생명은 풍부한 데이터와 데이터의 품질이다. 정밀한 구축 공정과 여러 전문가의 힘으로 완성되어 높은 품질을 지니게 된 본 사업의 최종 산출물은 산업의 발전에 큰 역할을 할 것이 분명하다.

## 2 최종 산출물

### □ 최종 납품 결과물

본 사업을 진행하며 참여 기관인 데이터리와 한림대학교는 주관기관인 국립국어원에 개체명 분석 웹 말뭉치 500만 어절과 수정 및 보완이 이루어진 지침, 또 사업의 전체적인 진행 내용을 기술한 최종 보고서를 납품하며 본 사업을 마무리한다.

산출물	내용	매체	시 기
개체명 분석 말뭉치(최종)	- 개체명 분석 말뭉치 * 전체 구축 분량(500만 어절)	휴대용 저장매체 3개	사업 종료 시
최종 보고서	- 개체명 분석 지침 - 사업 수행 내용 및 결과 요약 보고서	보고서 30부 (인쇄본)	

<표 15> 최종 납품물

## □ 개체명 말뭉치 분석 결과

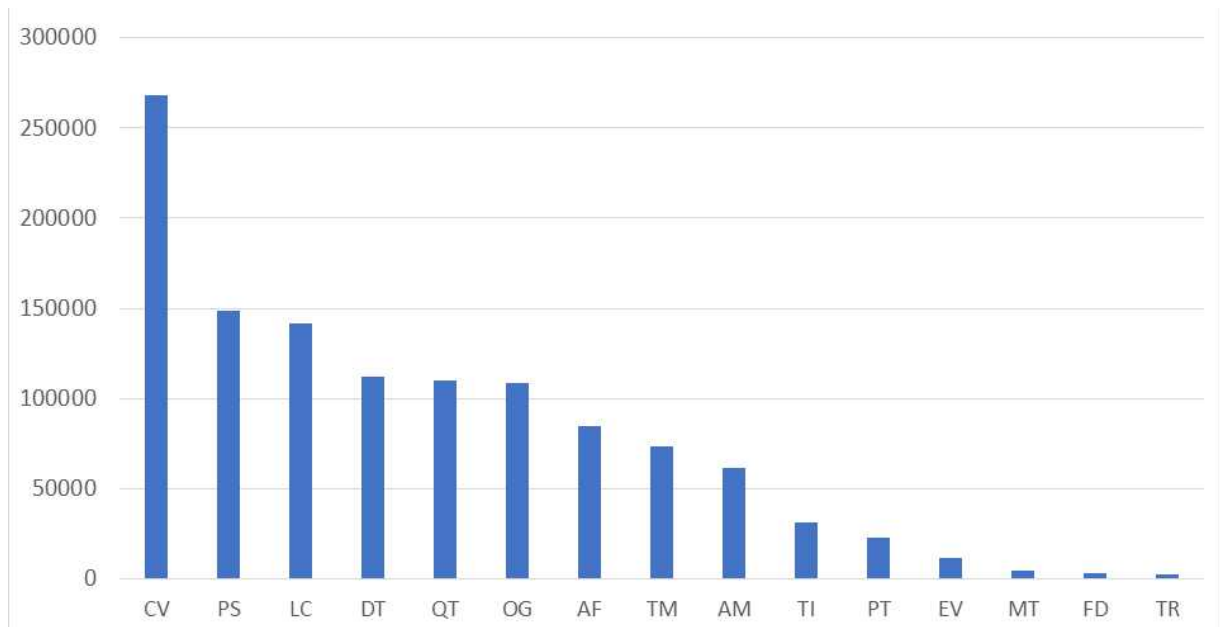
이번 사업의 개체명 분석 대상인 웹 원시 말뭉치는 <표 13>과 같은 JSON 형식으로 구성되어 있다. 분류에 따른 50개의 원시 말뭉치 JSON 파일이 있으며 파일 분류별로 블로그, 게시판, 누리소통망 등의 출처를 가진다. JSON 파일 내에는 분류 정보와 출처 문서의 정보가 포함되어 있고 document > sentence > form의 내용이 분석 대상이 되는 문장으로 총 1,354,881개이다.

```

{
  "id": "EXRW2006110001",
  "metadata": {
    "title": "EBXX1908-01",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2020",
    "category": "웹말뭉치>구축",
    "annotation_level": [
      "원시"
    ],
    "sampling": "부분 추출 - 임의 추출"
  },
  "document": [
    {
      "id": "EBRW1908000115.3",
      "metadata": {
        "title": "고양이 스노우(SNOW) 카메라 인식 스티커 / 고양이 사진찍기 / 아기 코삭 치즈 냅이 키우기 / 새끼고양이 키우기 / 오묘한 기록",
        "author": "맑고흐림",
        "publisher": "네이버 블로그",
        "date": "20190709",
        "crawl_date": "20190802 02:07:57",
        "url": "http://blog.naver.com/PostView.nhn?blogId=cloud_ent_&logNo=221581261106"
      },
      "sentence": [
        {
          "id": "EBRW1908000115.3.1.1",
          "form": "며칠전 고다(고양이라서 다행이야) 카페에 들어갔다가 재밌는 사진을 발견했다."
        }
      ]
    }
  ]
}

```

<표 16> 원시 말뭉치 JSON 형식



<표 17> 상위 15개 카테고리 출현 빈도

개체명 분석 웹 말뭉치 구축 결과 대분류별 출현 빈도는 위 그래프와 같다. 주로 의식주, 사회, 예술, 문화, 제도, 법과 관련된 CIVILIZATION이 약 268,144

개로 다른 대분류에 비해 월등히 많이 부착되었다. CIVILIZATION는 가장 많은 21개의 세부분류를 가지고 있고, 말뚝치 구축의 대상이 우리가 살아가는 사회와 생활을 가장 밀접하게 담고 있는 개개인의 웹 문서라는 점이 크게 작용한 듯 보인다.

다음으로는 인명, 별명, 캐릭터명, 애완동물 이름과 관련된 PERSON이 약 148,395개 부착되었다. 주로 개인의 일상에서 등장하는 애완동물의 이름이나 지인의 이름, 또는 닉네임이 다수 등장하였으며 연예인, 만화 캐릭터, 가상의 인물 명칭도 많이 등장하였다.

다음으로는 지역, 장소, 지형, 지리 등과 관련된 LOCATION이 약 141,691개 부착되었다. 관광지과 명소, 식당이나 쇼핑몰, 그 외 공원과 해변 등이 다수 등장한 것으로 보아 이 역시 개인의 일상을 담은 웹 문서의 특성이 크게 작용한 결과로 보인다.

이외에 날짜를 나타내는 DATE 111,913개, 수량과 순서 등 기타 수사와 관련된 QUANTITY가 110,008개, 기관과 단체를 나타내는 ORGANIZATION이 108,660개로 각각 비슷한 빈도로 부착되었다. 인공물과 관련된 ARTIFACTS는 84,918개, 14개의 대분류에 포함되지 않는 나머지 개체명 대상인 색, 세포, 장치, 질병, IT 서비스 등을 포함하는 TERM은 73,542개, 동물과 신체 및 몸의 명칭과 관련된 ANIMAL은 61,486개로 앞선 대분류들에 비해 십만 단위에서 만 단위로 하락하였다. 시간을 나타내는 TIME은 31,069개, 식물과 관련된 PLANT가 22,740개, 사건 및 행사 등과 관련된 EVENT는 11,815개로 각각 1만 개 정도 차이를 보였으며 원소 및 화학물을 나타내는 MATERIAL은 4,356개, 학문 분야를 나타내는 STUDY\_FIELD는 3,011개, 특정 이론, 법칙, 원리 등과 관련된 THEORY는 2,527개로 가장 적게 부착되었다.

각 대분류에 속하는 세부분류별 출현 빈도는 아래의 <표 14>와 같다.

개체명의 분류별 부착 빈도는 구축 말뚝치의 종류에 따라 달라지므로 웹 원시 말뚝치를 대상으로 한 사업에서 나타난 통계라는 점에서 의의가 있다.

CIVILIZATION	빈도
CV_FOOD	161871
CV_DRINK	37370
CV_RELATION	32677
CV_OCCUPATION	15893
CV_CLOTHING	11735
CV_POSITION	10720
CV_ART	5406
CV_SPORTS	4260
CV_LANGUAGE	2867
CV_FOOD_STYLE	1809
CV_TRIBE	1319
CV_LAW	1270
CV_PRIZE	561
CV_POLICY	524
CV_SPORTS_INST	457
CV_BUILDING_TYPE	252
CV_CURRENCY	137
CV_SPORTS_POSITION	135
CV_FUNDS	127
CV_TAX	124
CV_CULTURE	37

DATE	빈도
DT_DAY	51013
DT_DURATION	23548
DT_OTHERS	13009
DT_SEASON	11138
DT_YEAR	10268
DT_MONTH	4282
DT_WEEK	2241
DT_DYNASTY	392
DT_GEOAGE	31

THEORY	빈도
TR_SOCIAL_SCIENCE	842
TR_MEDICINE	704
TR_SCIENCE	639
TR_ART	243
TR_HUMANITIES	137
TR_OTHERS	56

PERSON	빈도
PS_NAME	132802
PS_NAME	132802
PS_CHARACTER	16508

LOCATION	빈도
LCP_CITY	34201
LC_OTHERS	30505
LCP_COUNTY	29149
LCP_COUNTRY	18675
LCP_PROVINCE	11975
LCP_CAPITALCITY	8260
LCG_ISLAND	3503
LCG_MOUNTAIN	3502
LCG_RIVER	2425
LC_SPACE	1847
LCG_CONTINENT	1323
LCG_OCEAN	466
LCG_BAY	312

QUANTITY	빈도
QT_COUNT	38993
QT_ORDER	24488
QT_OTHERS	12474
QT_PRICE	10582
QT_MAN_COUNT	9139
QT_AGE	5968
QT_PERCENTAGE	4177
QT_LENGTH	2267
QT_WEIGHT	2190
QT_ADDRESS	1748
QT_VOLUME	991
QT_TEMPERATURE	486
QT_SPORTS	422
QT_PHONE	330
QT_SIZE	311
QT_ALBUM	150
QT_SPEED	70
QT_CHANNEL	27

ORGANIZATION	빈도
OGG_FOOD	51995
OGG_ECONOMY	34674
OGG_POLITICS	5847
OGG_OTHERS	4791
OGG_EDUCATION	3963
OGG_ART	3068
OGG_SPORTS	2698
OGG_HOTEL	2622
OGG_MEDIA	1871
OGG_MILITARY	888
OGG_RELIGION	804
OGG_MEDICINE	422
OGG_LAW	347
OGG_LIBRARY	152
OGG_SCIENCE	115

ARTIFACTS	빈도
AFW_OTHER_PRODUCTS	20278
AFA_VIDEO	16480
AF_TRANSPORT	11880
AF_BUILDING	6183
AFA_MUSIC	5996
AF_ROAD	5283
AFA_DOCUMENT	5176
AF_CULTURAL_ASSET	3306
AF_MUSICAL_INSTRUMENT	1443
AFA_PERFORMANCE	1333
AF_WEAPON	999
AFW_SERVICE_PRODUCTS	591
AFA_ART_CRAFT	312

TERM	빈도
TMI_SERVICE	16949
TMI_HW	13817
TM_DIRECTION	12618
TM_COLOR	8794
TMM_DISEASE	6733
TMI_SITE	4178
TM_CELL_TISSUE_ORGAN	3632
TMM_DRUG	1559
TM_SHAPE	1294
TMI_SW	1168
TMI_MODEL	711
TM_SPORTS	647
TMIG_GENRE	496
TMI_PROJECT	251
TMI_EMAIL	76
TM_CLIMATE	23

ANIMAL	빈도
AM_PART	43425
AM_MAMMALIA	17021
AM_OTHERS	2400
AM_BIRD	1725
AM_INSECT	1422
AM_FISH	1195
AM_TYPE	691
AM_REPTILIA	615
AM_AMPHIBIA	355

TIME	빈도
TI_DURATION	22439
TI_OTHERS	4769
TI_HOUR	4109
TI_MINUTE	407
TI_SECOND	146

PLANT	빈도
PT_FRUIT	5875
PT_GRASS	4820
PT_FLOWER	4442
PT_PART	4354
PT_TREE	1434
PT_OTHERS	1211
PT_TYPE	723

EVENT	빈도
EV_FESTIVAL	4980
EV_OTHERS	4771
EV_SPORTS	2030
EV_ACTIVITY	253
EV_WAR_REVOLUTION	250

MATERIAL	빈도
MT_CHEMICAL	3477
MT_METAL	503
MT_ROCK	371
MT_ELEMENT	313

STUDY_FIELD	빈도
FD_SCIENCE	1012
FD_HUMANITIES	834
FD_MEDICINE	563
FD_ART	296
FD_SOCIAL_SCIENCE	264
FD_OTHERS	28

<표 18> 세부분류 전체 빈도

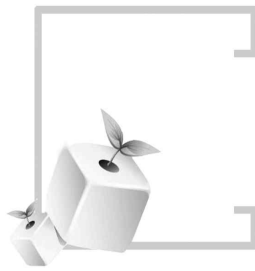
국립국어원이 주관한 ‘2020년 개체명 말뭉치 연구 분석’ 사업은 웹 원시 말뭉치라는 특수한 말뭉치를 대상으로 진행하였으며 그에 맞추어 지침의 변화와 사용 도구의 수정이 많이 이루어졌다. 사업이 성공적으로 마무리 되었지만 여전히 아쉬운 점은 남아 있다. 추후 진행될 관련 사업에 도움이 되고자 남은 문제에 대해 정리하고자 한다.

원시 말뭉치에 비해 주석(태깅) 말뭉치를 구축하는 사업은 상대적으로 일관성 있고 명확한 지침, 고성능·고효율 프로그램, 국어학 지식을 지니고 있으며 지침을 숙지한 작업자가 갖추어져야 성공적으로 완수될 수 있다. 이 중 가장 우선적으로 마련되어야 하는 것이 ‘말뭉치 구축 지침’이다. 본 사업은 말뭉치 구축과 지침 정비를 동시에 수행하도록 계획되어, 500만 어절의 개체명 말뭉치를 구축하면서 지침을 수정하는 작업이 병행되었다. 기존의 지침은 비교적 특수성이 없는 문어와 구어로 이루어진 말뭉치를 기준으로 만들어졌고 이번 사업의 대상인 웹 말뭉치에서 나타나는 특징인 문맥 파악 불가, 지나친 함축, 인터넷 용어, 줄임 표현 등을 포괄할 수 없는 부분이 있었다. 이로 인해 대대적인 수정과 변경이 이루어져 종료 시점까지 지침이 변경되어 태깅이 끝난 전체 데이터를 다시 수정해야 하는 작업이 수차례 반복되었다. 이는 작업자의 혼란뿐 아니라 기계 학습에도 오류를 야기했다. 향후의 개체명 말뭉치 구축 사업에서는 구축 대상 말뭉치의 특성을 고려하여 지침이 그 특성을 아우를 수 있는지 확인하고 보다 정비된, 완성도가 높은 지침이 배포된 상태로 말뭉치 구축이 시작된다면 더 빠르고 정확하게 작업이 수행될 것이다. 이를 위해 이번에 수정·보완된 개체명 말뭉치 구축 지침에 남아 있는 문제를 정리하고자 한다.

하나의 단어나 구가 둘 이상의 개체명에 속하는 경우가 있다. 한 가지는 체계적인 다의성을 가지는 경우이다. ‘서울시’가 OGG에도 속하고, LCP에도 속하는 것이 그러한 경우이다. 이때에는 문맥 우선 원칙을 적용하면 어떤 태그를 부착할지 결정할 수 있다. 다른 한 가지는 개체명 분류 자체가 서로 다른 층위의 속성을 가지는 경우이다. ‘해운대’는 해변이면서 관광지로 유명해진 곳이고, ‘남대문’은 본래 건축물인데 문화재로 지정된 것이다. 기존 지침에서는 전자는 LC\_TOUR로, 후자는 AF\_CUTURAL\_ASSET으로 태깅하였다. 그런데 LC\_TOUR의 경우 목록을 총망라할 수도 없고 작업자마다 판단이 다를 수 있다는 문제가 있어서 올해 지침에서는 LC\_TOUR를 삭제하였다. 그러나 상대적으로 목록이 한

정적인 문화재인 건축물은 여전히 AF\_CUTURAL\_ASSET으로 태깅하고 있다. 전자는 본래적인 속성에 따라 태깅하기로 한 것인데, 후자는 부차적인 속성에 따라 태깅하는 것이라 일관성이 없다는 문제가 있다. 향후 AF\_CUTURAL\_ASSET의 삭제에 대해서도 검토할 필요가 있다.

실용적인 목적을 위해서는 LC\_TOUR나 AF\_CUTURAL\_ASSET과 같은 항목이 필요할 수도 있다. 유일 태깅 원칙이라는 대원칙을 지키면서는 이 문제를 해결하기 어렵다. 기본적인 개체명 태깅은 현재 지침에 따라 하도록 하되 필요한 경우 부차적인 속성을 표시하는 태그를 추가하여 중복 태깅할 수 있게 하는 방안도 검토할 만하다. 예를 들어 경리단길은 ‘AF\_ROAD:TOUR’와 같이 태깅하고 남대문은 ‘AF\_BUILDING:CUTURAL\_ASSET’과 같이 태깅하는 것이다. 이를 위해서는 부차적인 태그에 대한 수요 조사와 태그 체계에 대한 정밀한 연구가 필요하다.



## 부록

# 2020년 개체명 분석 말뭉치 구축 최종 지침







# 2020 개체명 분석 말뭉치 구축 지침

---

ver. 1.6.

---

# 2020 개체명 분석 말뭉치 구축 지침

---

ver. 1.6.

2020.12.19.

---

## 가. 개요

## 나. 태깅 대상

## 다. 태깅 단위

1. 태깅 단위 결정의 원리 - 고유명
  - 1.1. 최장 단위 태깅(원칙)
  - 1.2. 최소 단위 태깅(예외)
2. 태깅 단위 결정의 원리 - 보통명사
  - 2.1. 최장 단위 태깅(원칙)
  - 2.2. 최소 단위 태깅(예외)
  - 2.3. 주의 사항

## 라. 태깅 범주

1. 태깅 범주 결정의 대원칙 - 유일 태깅 원칙
2. 태깅 범주 결정의 원리
  - 2.1. 문맥 우선 적용(원칙)
  - 2.2. 우선순위 적용(예외)

## 마. 특수 항목 태깅

1. 괄호 안의 원어 정보
2. 괄호 안의 부가 정보
3. 원어절 사이에 기호가 삽입되어 있는 경우
4. 날짜/시간/수량 N + 보조사/접사/단위성 의존명사
5. 영어나 한자 등의 외국어
6. 줄임말
7. 개체명의 일부로 해당 개체를 가리키는 경우
8. 비유적 표현
9. 개체명과 조사/어미가 하나의 음절로 실현된 경우
10. 오자나 오타

## 바. 세부분류 개체명 정의

### 1. PERSON

- 1.1. PS\_NAME
  - 1.1.1. 정의 및 설명
  - 1.1.2. 분석 대상
  - 1.1.3. 예문

- 1.1.4. 주의 사항
- 1.2. PS\_CHARACTER
  - 1.2.1. 정의 및 설명
  - 1.2.2. 분석 대상
  - 1.2.3. 예문
  - 1.2.4. 주의 사항
- 1.3. PS\_PET
  - 1.3.1. 정의 및 설명
  - 1.3.2. 분석 대상
  - 1.3.3. 예문
  - 1.3.4. 주의 사항

## **2. STUDY\_FIELD**

- 2.1. FD\_SCIENCE
  - 2.1.1. 정의 및 설명
  - 2.1.2. 분석 대상
  - 2.1.3. 예문
  - 2.1.4. 주의 사항
- 2.2. FD\_SOCIAL\_SCIENCE
  - 2.2.1. 정의 및 설명
  - 2.2.2. 분석 대상
  - 2.2.3. 예문
  - 2.2.4. 주의 사항
- 2.3. FD\_MEDICINE
  - 2.3.1. 정의 및 설명
  - 2.3.2. 분석 대상
  - 2.3.3. 예문
  - 2.3.4. 주의 사항
- 2.4. FD\_ART
  - 2.4.1. 정의 및 설명
  - 2.4.2. 분석 대상
  - 2.4.3. 예문
  - 2.4.4. 주의 사항
- 2.5. FD\_HUMANITIES
  - 2.5.1. 정의 및 설명
  - 2.5.2. 분석 대상
  - 2.5.3. 예문

2.5.4. 주의 사항

2.6. FD\_OTHERS

2.6.1. 정의 및 설명

2.6.2. 분석 대상

2.6.3. 예문

**3. THEORY**

3.1. TR\_SCIENCE

3.1.1. 정의 및 설명

3.1.2. 분석 대상

3.1.3. 예문

3.1.4. 주의 사항

3.2. TR\_SOCIAL\_SCIENCE

3.2.1. 정의 및 설명

3.2.2. 분석 대상

3.2.3. 예문

3.2.4. 주의 사항

3.3. TR\_MEDICINE

3.3.1. 정의 및 설명

3.3.2. 분석 대상

3.3.3. 예문

3.4. TR\_ART

3.4.1. 정의 및 설명

3.4.2. 분석 대상

3.4.3. 예문

3.4.4. 주의 사항

3.5. TR\_HUMANITIES

3.5.1. 정의 및 설명

3.5.2. 분석 대상

3.5.3. 예문

3.5.4. 주의 사항

3.6. TR\_OTHERS

3.6.1. 정의 및 설명

3.6.2. 분석 대상

3.6.3. 예문

3.6.4. 주의 사항

**4. ARTIFACTS**

- 4.1. AF\_BUILDING
  - 4.1.1. 정의 및 설명
  - 4.1.2. 분석 대상
  - 4.1.3. 예문
  - 4.1.4. 주의 사항
- 4.2. AF\_CULTURAL\_ASSET
  - 4.2.1. 정의 및 설명
  - 4.2.2. 분석 대상
  - 4.2.3. 예문
  - 4.2.4. 주의 사항
- 4.3. AF\_ROAD
  - 4.3.1. 정의 및 설명
  - 4.3.2. 분석 대상
  - 4.3.3. 예문
  - 4.3.4. 주의 사항
- 4.4. AF\_TRANSPORT
  - 4.4.1. 정의 및 설명
  - 4.4.2. 분석 대상
  - 4.4.3. 예문
  - 4.4.4. 주의 사항
- 4.5. AF\_MUSICAL\_INSTRUMENT
  - 4.5.1. 정의 및 설명
  - 4.5.2. 분석 대상
  - 4.5.3. 예문
  - 4.5.4. 주의 사항
- 4.6. AF\_WEAPON
  - 4.6.1. 정의 및 설명
  - 4.6.2. 분석 대상
  - 4.6.3. 예문
  - 4.6.4. 주의 사항
- 4.7. AF\_ART\_WORKS
  - 4.7.1. AFA\_DOCUMENT
    - 4.7.1.1. 정의 및 설명
    - 4.7.1.2. 분석 대상
    - 4.7.1.3. 예문
    - 4.7.1.4. 주의 사항

#### 4.7.2. AFA\_PERFORMANCE

4.7.2.1. 정의 및 설명

4.7.2.2. 분석 대상

4.7.2.3. 예문

4.7.2.4. 주의 사항

#### 4.7.3. AFA\_VIDEO

4.7.3.1. 정의 및 설명

4.7.3.2. 분석 대상

4.7.3.3. 예문

4.7.3.4. 주의 사항

#### 4.7.4. AFA\_ART\_CRAFT

4.7.4.1. 정의 및 설명

4.7.4.2. 분석 대상

4.7.4.3. 예문

4.7.4.4. 주의 사항

#### 4.7.5. AFA\_MUSIC

4.7.5.1. 정의 및 설명

4.7.5.2. 분석 대상

4.7.5.3. 예문

4.7.5.4. 주의 사항

#### 4.8. AF\_WARES

##### 4.8.1. AFW\_SERVICE\_PRODUCTS

4.8.1.1. 정의 및 설명

4.8.1.2. 분석 대상

4.8.1.3. 예문

4.8.1.4. 주의 사항

##### 4.8.2. AFW\_OTHER\_PRODUCTS

4.8.2.1. 정의 및 설명

4.8.2.2. 분석 대상

4.8.2.3. 예문

4.8.2.4. 주의 사항

### 5. ORGANIZATION

#### 5.1. OGG\_ECONOMY

5.1.1. 정의 및 설명

5.1.2. 분석 대상

5.1.3. 예문

- 5.1.4. 주의 사항
- 5.2. OGG\_EDUCATION
  - 5.2.1. 정의 및 설명
  - 5.2.2. 분석 대상
  - 5.2.3. 예문
  - 5.2.4. 주의 사항
- 5.3. OGG\_MILITARY
  - 5.3.1. 정의 및 설명
  - 5.3.2. 분석 대상
  - 5.3.3. 예문
  - 5.3.4. 주의 사항
- 5.4. OGG\_MEDIA
  - 5.4.1. 정의 및 설명
  - 5.4.2. 분석 대상
  - 5.4.3. 예문
  - 5.4.4. 주의 사항
- 5.5. OGG\_SPORTS
  - 5.5.1. 정의 및 설명
  - 5.5.2. 분석 대상
  - 5.5.3. 예문
  - 5.5.4. 주의 사항
- 5.6. OGG\_ART
  - 5.6.1. 정의 및 설명
  - 5.6.2. 분석 대상
  - 5.6.3. 예문
  - 5.6.4. 주의 사항
- 5.7. OGG\_MEDICINE
  - 5.7.1. 정의 및 설명
  - 5.7.2. 분석 대상
  - 5.7.3. 예문
  - 5.7.4. 주의 사항
- 5.8. OGG\_RELIGION
  - 5.8.1. 정의 및 설명
  - 5.8.2. 분석 대상
  - 5.8.3. 예문
  - 5.8.4. 주의 사항



- 5.9. OGG\_SCIENCE
  - 5.9.1. 정의 및 설명
  - 5.9.2. 분석 대상
  - 5.9.3. 예문
  - 5.9.4. 주의 사항
- 5.10. OGG\_LIBRARY
  - 5.10.1. 정의 및 설명
  - 5.10.2. 분석 대상
  - 5.10.3. 예문
  - 5.10.4. 주의 사항
- 5.11. OGG\_LAW
  - 5.11.1. 정의 및 설명
  - 5.11.2. 분석 대상
  - 5.11.3. 예문
  - 5.11.4. 주의 사항
- 5.12. OGG\_POLITICS
  - 5.12.1. 정의 및 설명
  - 5.12.2. 분석 대상
  - 5.12.3. 예문
  - 5.12.4. 주의 사항
- 5.13. OGG\_FOOD
  - 5.13.1. 정의 및 설명
  - 5.13.2. 분석 대상
  - 5.13.3. 예문
  - 5.13.4. 주의 사항
- 5.14. OGG\_HOTEL
  - 5.14.1. 정의 및 설명
  - 5.14.2. 분석 대상
  - 5.14.3. 예문
  - 5.14.4. 주의 사항
- 5.15. OGG\_OTHERS
  - 5.15.1. 정의 및 설명
  - 5.15.2. 분석 대상
  - 5.15.3. 예문
  - 5.15.4. 주의 사항

## **6. LOCATION**

- 6.1. LCP\_COUNTRY
  - 6.1.1. 정의 및 설명
  - 6.1.2. 분석 대상
  - 6.1.3. 예문
  - 6.1.4. 주의 사항
- 6.2. LCP\_PROVINCE
  - 6.2.1. 정의 및 설명
  - 6.2.2. 분석 대상
  - 6.2.3. 예문
  - 6.2.4. 주의 사항
- 6.3. LCP\_COUNTY
  - 6.3.1. 정의 및 설명
  - 6.3.2. 분석 대상
  - 6.3.3. 예문
  - 6.3.4. 주의 사항
- 6.4. LCP\_CITY
  - 6.4.1. 정의 및 설명
  - 6.4.2. 분석 대상
  - 6.4.3. 예문
  - 6.4.4. 주의 사항
- 6.5. LCP\_CAPITALCITY
  - 6.5.1. 정의 및 설명
  - 6.5.2. 분석 대상
  - 6.5.3. 예문
  - 6.5.4. 주의 사항
- 6.6. LCG\_RIVER
  - 6.6.1. 정의 및 설명
  - 6.6.2. 분석 대상
  - 6.6.3. 예문
  - 6.6.4. 주의 사항
- 6.7. LCG\_OCEAN
  - 6.7.1. 정의 및 설명
  - 6.7.2. 분석 대상
  - 6.7.3. 예문
  - 6.7.4. 주의 사항
- 6.8. LCG\_BAY

- 6.8.1. 정의 및 설명
- 6.8.2. 분석 대상
- 6.8.3. 예문
- 6.8.4. 주의 사항
- 6.9. LCG\_MOUNTAIN
  - 6.9.1. 정의 및 설명
  - 6.9.2. 분석 대상
  - 6.9.3. 예문
  - 6.9.4. 주의 사항
- 6.10. LCG\_ISLAND
  - 6.10.1. 정의 및 설명
  - 6.10.2. 분석 대상
  - 6.10.3. 예문
  - 6.10.4. 주의 사항
- 6.11. LCG\_CONTINENT
  - 6.11.1. 정의 및 설명
  - 6.11.2. 분석 대상
  - 6.11.3. 예문
  - 6.11.4. 주의 사항
- 6.12. LC\_SPACE
  - 6.12.1. 정의 및 설명
  - 6.12.2. 분석 대상
  - 6.12.3. 예문
  - 6.12.4. 주의 사항
- 6.13. LC\_OTHERS
  - 6.13.1. 정의 및 설명
  - 6.13.2. 분석 대상
  - 6.13.3. 예문
  - 6.13.4. 주의 사항

## **7. CIVILIZATION**

- 7.1. CV\_CULTURE
  - 7.1.1. 정의 및 설명
  - 7.1.2. 분석 대상
  - 7.1.3. 예문
- 7.2. CV\_TRIBE
  - 7.2.1. 정의 및 설명

- 7.2.2. 분석 대상
- 7.2.3. 예문
- 7.2.4. 주의 사항
- 7.3. CV\_LANGUAGE
  - 7.3.1. 정의 및 설명
  - 7.3.2. 분석 대상
  - 7.3.3. 예문
  - 7.3.4. 주의 사항
- 7.4. CV\_POLICY
  - 7.4.1. 정의 및 설명
  - 7.4.2. 분석 대상
  - 7.4.3. 예문
  - 7.4.4. 주의 사항
- 7.5. CV\_LAW
  - 7.5.1. 정의 및 설명
  - 7.5.2. 분석 대상
  - 7.5.3. 예문
  - 7.5.4. 주의 사항
- 7.6. CV\_CURRENCY
  - 7.6.1. 정의 및 설명
  - 7.6.2. 분석 대상
  - 7.6.3. 예문
- 7.7. CV\_TAX
  - 7.7.1. 정의 및 설명
  - 7.7.2. 분석 대상
  - 7.7.3. 예문
- 7.8. CV\_FUNDS
  - 7.8.1. 정의 및 설명
  - 7.8.2. 분석 대상
  - 7.8.3. 예문
  - 7.8.4. 주의 사항
- 7.9. CV\_ART
  - 7.9.1. 정의 및 설명
  - 7.9.2. 분석 대상
  - 7.9.3. 예문
  - 7.9.4. 주의 사항

- 7.10. CV\_SPORTS
  - 7.10.1. 정의 및 설명
  - 7.10.2. 분석 대상
  - 7.10.3. 예문
  - 7.10.4. 주의 사항
- 7.11. CV\_SPORTS\_POSITION
  - 7.11.1. 정의 및 설명
  - 7.11.2. 분석 대상
  - 7.11.3. 예문
  - 7.11.4. 주의 사항
- 7.12. CV\_SPORTS\_INST
  - 7.12.1. 정의 및 설명
  - 7.12.2. 분석 대상
  - 7.12.3. 예문
  - 7.12.4. 주의 사항
- 7.13. CV\_PRIZE
  - 7.13.1. 정의 및 설명
  - 7.13.2. 분석 대상
  - 7.13.3. 예문
  - 7.13.4. 주의 사항
- 7.14. CV\_RELATION
  - 7.14.1. 정의 및 설명
  - 7.14.2. 분석 대상
  - 7.14.3. 예문
  - 7.14.4. 주의 사항
- 7.15. CV\_OCCUPATION
  - 7.15.1. 정의 및 설명
  - 7.15.2. 분석 대상
  - 7.15.3. 예문
  - 7.15.4. 주의 사항
- 7.16. CV\_POSITION
  - 7.16.1. 정의 및 설명
  - 7.16.2. 분석 대상
  - 7.16.3. 예문
  - 7.16.4. 주의 사항
- 7.17. CV\_FOOD

- 7.17.1. 정의 및 설명
- 7.17.2. 분석 대상
- 7.17.3. 예문
- 7.17.4. 주의 사항
- 7.18. CV\_DRINK
  - 7.18.1. 정의 및 설명
  - 7.18.2. 분석 대상
  - 7.18.3. 예문
  - 7.18.4. 주의 사항
- 7.19. CV\_FOOD\_STYLE
  - 7.19.1. 정의 및 설명
  - 7.19.2. 분석 대상
  - 7.19.3. 예문
  - 7.19.4. 주의 사항
- 7.20. CV\_CLOTHING
  - 7.20.1. 정의 및 설명
  - 7.20.2. 분석 대상
  - 7.20.3. 예문
  - 7.20.4. 주의 사항
- 7.21. CV\_BUILDING\_TYPE
  - 7.21.1. 정의 및 설명
  - 7.21.2. 분석 대상
  - 7.21.3. 예문

## **8. DATE**

- 8.1. DT\_DURATION
  - 8.1.1. 정의 및 설명
  - 8.1.2. 분석 대상
  - 8.1.3. 예문
  - 8.1.4. 주의 사항
- 8.2. DT\_DAY
  - 8.2.1. 정의 및 설명
  - 8.2.2. 분석 대상
  - 8.2.3. 예문
  - 8.2.4. 주의 사항
- 8.3. DT\_WEEK
  - 8.4.1. 정의 및 설명

- 8.4.2. 분석 대상
- 8.4.3. 예문
- 8.4.4. 주의 사항
- 8.4. DT\_MONTH
  - 8.4.1. 정의 및 설명
  - 8.4.2. 분석 대상
  - 8.4.3. 예문
  - 8.4.4. 주의 사항
- 8.5. DT\_YEAR
  - 8.5.1. 정의 및 설명
  - 8.5.2. 분석 대상
  - 8.5.3. 예문
  - 8.5.4. 주의 사항
- 8.6. DT\_SEASON
  - 8.6.1. 정의 및 설명
  - 8.6.2. 분석 대상
  - 8.6.3. 예문
  - 8.6.4. 주의 사항
- 8.7. DT\_GEOAGE
  - 8.7.1. 정의 및 설명
  - 8.7.2. 분석 대상
  - 8.7.3. 예문
  - 8.7.4. 주의 사항
- 8.8. DT\_DYNASTY
  - 8.8.1. 정의 및 설명
  - 8.8.2. 분석 대상
  - 8.8.3. 예문
  - 8.8.4. 주의 사항
- 8.9. DT\_OTHERS
  - 8.9.1. 정의 및 설명
  - 8.9.2. 분석 대상
  - 8.9.3. 예문
  - 8.9.4. 주의 사항

## **9. TIME**

- 9.1. TI\_DURATION
  - 9.1.1. 정의 및 설명

9.1.2. 분석 대상

9.1.3. 예문

9.1.4. 주의 사항

9.2. TI\_HOUR

9.2.1. 정의 및 설명

9.2.2. 분석 대상

9.2.3. 예문

9.2.4. 주의 사항

9.3. TI\_MINUTE

9.3.1. 정의 및 설명

9.3.2. 분석 대상

9.3.3. 예문

9.3.4. 주의 사항

9.4. TI\_SECOND

9.4.1. 정의 및 설명

9.4.2. 분석 대상

9.4.3. 예문

9.4.4. 주의 사항

9.5. TI\_OTHERS

9.5.1. 정의 및 설명

9.5.2. 분석 대상

9.5.3. 예문

9.5.4. 주의 사항

**10. QUANTITY**

10.1. QT\_AGE

10.1.1. 정의 및 설명

10.1.2. 분석 대상

10.1.3. 예문

10.1.4. 주의 사항

10.2. QT\_SIZE

10.2.1. 정의 및 설명

10.2.2. 분석 대상

10.2.3. 예문

10.2.4. 주의 사항

10.3. QT\_LENGTH

10.3.1. 정의 및 설명



- 10.3.2. 분석 대상
- 10.3.3. 예문
- 10.3.4. 주의 사항
- 10.4. QT\_COUNT
  - 10.4.1. 정의 및 설명
  - 10.4.2. 분석 대상
  - 10.4.3. 예문
  - 10.4.4. 주의 사항
- 10.5. QT\_MAN\_COUNT
  - 10.5.1. 정의 및 설명
  - 10.5.2. 분석 대상
  - 10.5.3. 예문
  - 10.5.4. 주의 사항
- 10.6. QT\_WEIGHT
  - 10.6.1. 정의 및 설명
  - 10.6.2. 분석 대상
  - 10.6.3. 예문
  - 10.6.4. 주의 사항
- 10.7. QT\_PERCENTAGE
  - 10.7.1. 정의 및 설명
  - 10.7.2. 분석 대상
  - 10.7.3. 예문
  - 10.7.4. 주의 사항
- 10.8. QT\_SPEED
  - 10.8.1. 정의 및 설명
  - 10.8.2. 분석 대상
  - 10.8.3. 예문
  - 10.8.4. 주의 사항
- 10.9. QT\_TEMPERATURE
  - 10.9.1. 정의 및 설명
  - 10.9.2. 분석 대상
  - 10.9.3. 예문
  - 10.9.4. 주의 사항
- 10.10. QT\_VOLUME
  - 10.10.1. 정의 및 설명
  - 10.10.2. 분석 대상

- 10.10.3. 예문
- 10.10.4. 주의 사항
- 10.11. QT\_ORDER
  - 10.11.1. 정의 및 설명
  - 10.11.2. 분석 대상
  - 10.11.3. 예문
  - 10.11.4. 주의 사항
- 10.12. QT\_PRICE
  - 10.12.1. 정의 및 설명
  - 10.12.2. 분석 대상
  - 10.12.3. 예문
  - 10.12.4. 주의 사항
- 10.13. QT\_PHONE
  - 10.13.1. 정의 및 설명
  - 10.13.2. 분석 대상
  - 10.13.3. 예문
- 10.14. QT\_SPORTS
  - 10.14.1. 정의 및 설명
  - 10.14.2. 분석 대상
  - 10.14.3. 예문
  - 10.14.4. 주의 사항
- 10.15. QT\_CHANNEL
  - 10.15.1. 정의 및 설명
  - 10.15.2. 분석 대상
  - 10.15.3. 예문
  - 10.15.4. 주의 사항
- 10.16. QT\_ALBUM
  - 10.16.1. 정의 및 설명
  - 10.16.2. 분석 대상
  - 10.16.3. 예문
  - 10.16.4. 주의 사항
- 10.17. QT\_ADDRESS
  - 10.17.1. 정의 및 설명
  - 10.17.2. 분석 대상
  - 10.17.3. 예문
  - 10.17.4. 주의 사항

10.18. QT\_OTHERS

10.18.1. 정의 및 설명

10.18.2. 분석 대상

10.18.3. 예문

10.18.4. 주의 사항

**11. EVENT**

11.1. EV\_ACTIVITY

11.1.1. 정의 및 설명

11.1.2. 분석 대상

11.1.3. 예문

11.1.4. 주의 사항

11.2. EV\_WAR\_REVOLUTION

11.2.1. 정의 및 설명

11.2.2. 분석 대상

11.2.3. 예문

11.2.4. 주의 사항

11.3. EV\_SPORTS

11.3.1. 정의 및 설명

11.3.2. 분석 대상

11.3.3. 예문

11.3.4. 주의 사항

11.4. EV\_FESTIVAL

11.4.1. 정의 및 설명

11.4.2. 분석 대상

11.4.3. 예문

11.4.4. 주의 사항

11.5. EV\_OTHERS

11.5.1. 정의 및 설명

11.5.2. 분석 대상

11.5.3. 예문

11.5.4. 주의 사항

**12. ANIMAL**

12.1. AM\_INSECT

12.1.1. 정의 및 설명

12.1.2. 분석 대상

12.1.3. 예문

- 12.2. AM\_BIRD
  - 12.2.1. 정의 및 설명
  - 12.2.2. 분석 대상
  - 12.2.3. 예문
  - 12.2.4. 주의 사항
- 12.3. AM\_FISH
  - 12.3.1. 정의 및 설명
  - 12.3.2. 분석 대상
  - 12.3.3. 예문
  - 12.3.4. 주의 사항
- 12.4. AM\_MAMMALIA
  - 12.4.1. 정의 및 설명
  - 12.4.2. 분석 대상
  - 12.4.3. 예문
  - 12.4.4. 주의 사항
- 12.5. AM\_AMPHIBIA
  - 12.5.1. 정의 및 설명
  - 12.5.2. 분석 대상
  - 12.5.3. 예문
- 12.6. AM\_REPTILIA
  - 12.6.1. 정의 및 설명
  - 12.6.2. 분석 대상
  - 12.6.3. 예문
- 12.7. AM\_TYPE
  - 12.7.1. 정의 및 설명
  - 12.7.2. 분석 대상
  - 12.7.3. 예문
  - 12.7.4. 주의 사항
- 12.8. AM\_PART
  - 12.8.1. 정의 및 설명
  - 12.8.2. 분석 대상
  - 12.8.3. 예문
  - 12.8.4. 주의 사항
- 12.9. AM\_OTHERS
  - 12.9.1. 정의 및 설명
  - 12.9.2. 분석 대상

12.9.3. 예문

12.9.4. 주의 사항

### **13. PLANT**

#### 13.1. PT\_FRUIT

13.1.1. 정의 및 설명

13.1.2. 분석 대상

13.1.3. 예문

13.1.4. 주의 사항

#### 13.2. PT\_FLOWER

13.2.1. 정의 및 설명

13.2.2. 분석 대상

13.2.3. 예문

13.2.4. 주의 사항

#### 13.3. PT\_TREE

13.3.1. 정의 및 설명

13.3.2. 분석 대상

13.3.3. 예문

13.3.4. 주의 사항

#### 13.4. PT\_GRASS

13.4.1. 정의 및 설명

13.4.2. 분석 대상

13.4.3. 예문

13.4.4. 주의 사항

#### 13.5. PT\_TYPE

13.5.1. 정의 및 설명

13.5.2. 분석 대상

13.5.3. 예문

13.5.4. 주의 사항

#### 13.6. PT\_PART

13.6.1. 정의 및 설명

13.6.2. 분석 대상

13.6.3. 예문

13.6.4. 주의 사항

#### 13.7. PT\_OTHERS

13.7.1. 정의 및 설명

13.7.2. 분석 대상

13.7.3. 예문

13.7.4. 주의 사항

## **14. MATERIAL**

### **14.1. MT\_ELEMENT**

14.1.1. 정의 및 설명

14.1.2. 분석 대상

14.1.3. 예문

14.1.4. 주의 사항

### **14.2. MT\_METAL**

14.2.1. 정의 및 설명

14.2.2. 분석 대상

14.2.3. 예문

14.2.4. 주의 사항

### **14.3. MT\_ROCK**

14.3.1. 정의 및 설명

14.3.2. 분석 대상

14.3.3. 예문

14.3.4. 주의 사항

### **14.4. MT\_CHEMICAL**

14.4.1. 정의 및 설명

14.4.2. 분석 대상

14.4.3. 예문

14.4.4. 주의 사항

## **15. TERM**

### **15.1. TM\_COLOR**

15.1.1. 정의 및 설명

15.1.2. 분석 대상

15.1.3. 예문

15.1.4. 주의 사항

### **15.2. TM\_DIRECTION**

15.2.1. 정의 및 설명

15.2.2. 분석 대상

15.2.3. 예문

15.2.4. 주의 사항

### **15.3. TM\_CLIMATE**

15.3.1. 정의 및 설명

- 15.3.2. 분석 대상
- 15.3.3. 예문
- 15.3.4. 주의 사항
- 15.4. TM\_SHAPE
  - 15.4.1. 정의 및 설명
  - 15.4.2. 분석 대상
  - 15.4.3. 예문
  - 15.4.4. 주의 사항
- 15.5. TM\_CELL\_TISSUE\_ORGAN
  - 15.5.1. 정의 및 설명
  - 15.5.2. 분석 대상
  - 15.5.3. 예문
  - 15.5.4. 주의 사항
- 15.6. TMM\_DISEASE
  - 15.6.1. 정의 및 설명
  - 15.6.2. 분석 대상
  - 15.6.3. 예문
  - 15.6.4. 주의 사항
- 15.7. TMM\_DRUG
  - 15.7.1. 정의 및 설명
  - 15.7.2. 분석 대상
  - 15.7.3. 예문
  - 15.7.4. 주의 사항
- 15.8. TMI\_HW
  - 15.8.1. 정의 및 설명
  - 15.8.2. 분석 대상
  - 15.8.3. 예문
  - 15.8.4. 주의 사항
- 15.9. TMI\_SW
  - 15.9.1. 정의 및 설명
  - 15.9.2. 분석 대상
  - 15.9.3. 예문
  - 15.9.4. 주의 사항
- 15.10. TMI\_SITE
  - 15.10.1. 정의 및 설명
  - 15.10.2. 분석 대상

- 15.10.3. 예문
- 15.10.4. 주의 사항
- 15.11. TMI\_EMAIL
  - 15.11.1. 정의 및 설명
  - 15.11.2. 분석 대상
  - 15.11.3. 예문
- 15.12. TMI\_MODEL
  - 15.12.1. 정의 및 설명
  - 15.12.2. 분석 대상
  - 15.12.3. 예문
  - 15.12.4. 주의 사항
- 15.13. TMI\_SERVICE
  - 15.13.1. 정의 및 설명
  - 15.13.2. 분석 대상
  - 15.13.3. 예문
  - 15.13.4. 주의 사항
- 15.14. TMI\_PROJECT
  - 15.14.1. 정의 및 설명
  - 15.14.2. 분석 대상
  - 15.14.3. 예문
  - 15.14.4. 주의 사항
- 15.15. TMIG\_GENRE
  - 15.15.1. 정의 및 설명
  - 15.15.2. 분석 대상
  - 15.15.3. 예문
  - 15.15.4. 주의 사항
- 15.16. TM\_SPORTS
  - 15.16.1. 정의 및 설명
  - 15.16.2. 분석 대상
  - 15.16.3. 예문
  - 15.16.4. 주의 사항



## 가. 개요

이 문서는 개체명 분석 말뭉치 구축 지침으로, 각 개체명 부류의 정의 및 설명, 분석 대상, 예문, 주의 사항을 기술하고 있다.

- 개체명 태그 세트는 15개의 대분류와 145개의 소분류, 7개의 세분류로 구성되어 있다. 최말단 태그로 하면 총 150개의 세부분류로 구분되어 있다.
- 개체명의 대분류는 15가지로, PERSON(인명), STUDY\_FIELD(학문 분야 및 학파), THEORY(이론), ARTIFACTS(인공물), ORGANIZATION(기관 및 조직), LOCATION(지명), CIVILIZATION(문명), DATE(날짜), TIME(시간), QUANTITY(수량), EVENT(사건 및 사고), ANIMAL(동물), PLANT(식물), MATERIAL(물질), TERM(기타 용어)으로 이루어져 있다.
- 15개의 대분류는 각각 여러 개의 하위부류로 세분화되어 있는데, 목적에 따라 대분류만 활용해도 되고 세부분류까지 활용해도 된다. 이 지침은 일종의 가이드라인의 성격을 갖는 것으로, 목적에 따라 일부 부류를 사용하지 않거나 필요한 부류를 더 추가할 수 있다.
- 개체명의 대분류, 소분류, 세분류의 목록을 제시하면 아래 표와 같다.

대분류	소분류	세분류
1. PERSON(PS)	PS_NAME	
	PS_CHARACTER	
	PS_PET	
2. STUDY_FIELD(FD)	FD_SCIENCE	
	FD_SOCIAL_SCIENCE	
	FD_MEDICINE	
	FD_ART	
	FD_HUMANITIES	
	FD_OTHERS	
3. THEORY(TR)	TR_SCIENCE	
	TR_SOCIAL_SCIENCE	
	TR_MEDICINE	
	TR_ART	
	TR_HUMANITIES	
	TR_OTHERS	
4. ARTIFACTS(AF)	AF_BUILDING	
	AF_CULTURAL_ASSET	
	AF_ROAD	
	AF_TRANSPORT	
	AF_MUSICAL_INSTRUMENT	

	AF_WEAPON	
	AF_ART_WORKS	AFA_DOCUMENT
		AFA_PERFORMANCE
		AFA_VIDEO
		AFA_ART_CRAFT
		AFA_MUSIC
	AF_WARES	AFW_SERVICE_PRODUCTS
<b>5. ORGANIZATION(OG)</b>		AFW_OTHER_PRODUCTS
	OGG_ECONOMY	
	OGG_EDUCATION	
	OGG_MILITARY	
	OGG_MEDIA	
	OGG_SPORTS	
	OGG_ART	
	OGG_MEDICINE	
	OGG_RELIGION	
	OGG_SCIENCE	
	OGG_LIBRARY	
	OGG_LAW	
	OGG_POLITICS	
	OGG_FOOD	
	OGG_HOTEL	
	OGG_OTHERS	
<b>6. LOCATION(LC)</b>	LCP_COUNTRY	
	LCP_PROVINCE	
	LCP_COUNTY	
	LCP_CITY	
	LCP_CAPITALCITY	
	LCG_RIVER	
	LCG_OCEAN	
	LCG_BAY	
	LCG_MOUNTAIN	
	LCG_ISLAND	
	LCG_CONTINENT	
	LC_SPACE	
	LC_OTHERS	
<b>7. CIVILIZATION(CV)</b>	CV_CULTURE	

	CV_TRIBE	
	CV_LANGUAGE	
	CV_POLICY	
	CV_LAW	
	CV_CURRENCY	
	CV_TAX	
	CV_FUNDS	
	CV_ART	
	CV_SPORTS	
	CV_SPORTS_POSITION	
	CV_SPORTS_INST	
	CV_PRIZE	
	CV_RELATION	
	CV_OCCUPATION	
	CV_POSITION	
	CV_FOOD	
	CV_DRINK	
	CV_FOOD_STYLE	
	CV_CLOTHING	
	CV_BUILDING_TYPE	
<b>8. DATE(DT)</b>	DT_DURATION	
	DT_DAY	
	DT_WEEK	
	DT_MONTH	
	DT_YEAR	
	DT_SEASON	
	DT_GEOAGE	
	DT_DYNASTY	
	DT_OTHERS	
<b>9. TIME(TI)</b>	TI_DURATION	
	TI_HOUR	
	TI_MINUTE	
	TI_SECOND	
	TI_OTHERS	
<b>10. QUANTITY(QT)</b>	QT_AGE	
	QT_SIZE	
	QT_LENGTH	

	QT_COUNT	
	QT_MAN_COUNT	
	QT_WEIGHT	
	QT_PERCENTAGE	
	QT_SPEED	
	QT_TEMPERATURE	
	QT_VOLUME	
	QT_ORDER	
	QT_PRICE	
	QT_PHONE	
	QT_SPORTS	
	QT_CHANNEL	
	QT_ALBUM	
	QT_ADDRESS	
	QT_OTHERS	
<b>11. EVENT(EV)</b>	EV_ACTIVITY	
	EV_WAR_REVOLUTION	
	EV_SPORTS	
	EV_FESTIVAL	
	EV_OTHERS	
<b>12. ANIMAL(AM)</b>	AM_INSECT	
	AM_BIRD	
	AM_FISH	
	AM_MAMMALIA	
	AM_AMPHIBIA	
	AM_REPTILIA	
	AM_TYPE	
	AM_PART	
	AM_OTHERS	
<b>13. PLANT(PT)</b>	PT_FRUIT	
	PT_FLOWER	
	PT_TREE	
	PT_GRASS	
	PT_TYPE	
	PT_PART	
	PT_OTHERS	
<b>14. MATERIAL(MT)</b>	MT_ELEMENT	
	MT_METAL	

15. TERM(TM)	MT_ROCK	
	MT_CHEMICAL	
	TM_COLOR	
	TM_DIRECTION	
	TM_CLIMATE	
	TM_SHAPE	
	TM_CELL_TISSUE_ORGAN	
	TMM_DISEASE	
	TMM_DRUG	
	TMI_HW	
	TMI_SW	
	TMI_SITE	
	TMI_EMAIL	
	TMI_MODEL	
	TMI_SERVICE	
	TMI_PROJECT	
	TMIG_GENRE	
	TM_SPORTS	

분류 명칭	정의 및 예
<b>1. PERSON(PS)</b>	- 인명 및 인물의 별칭
1.1. PS_NAME	- 실존/비실존 인물의 이름, 별칭, 인터넷 사용자명 - 예) 흥선대원군, 아이유, 단군, 제우스, 돌부처, 썬양, sojujsh
1.2. PS_CHARACTER	- 가상 캐릭터의 이름, 작품 속 등장인물의 이름 - 예) 뽀로로, 루피, 심슨, 엘사, 어피치, 라이언, 무지, 해리 포터, 유시진, 제임스 본드
1.3. PS_PET	- 반려동물의 이름 - 예) 뽀빠, 쿠키, 탄이, 절미, 달리, 김또또, 이몽치
<b>2. STUDY_FIELD(FD)</b>	- 학문 분야
2.1. FD_SCIENCE	- 과학 학문 분야 및 학파(자연과학과 공학만으로 한정) - 예) 물리학, 생명과학, 수학, 기계공학, 수의학, 자동차공학, 전기공학
2.2. FD_SOCIAL_SCIENCE	- 사회과학 학문 분야 및 학파(정치/경제/사회와 관련된 분야) - 예) 인류학, 사회학, 경제학, 정치학, 심리학/고전학파, 신고전 학파
2.3. FD_MEDICINE	- 의학 관련 학문 분야 및 분과 - 예) 의학, 병리학, 보건의료학, 신경과학/비뇨기과, 신경과, 내과
2.4. FD_ART	- 예술 관련 학문 분야 및 학파

	- 예) 문예학, 음악사학, 미술사학, 공연예술학/인상파, 점묘파, 신경향파
2.5. FD_HUMANITIES	- 인문학 관련 학문 분야 및 학파 - 예) 언어학, 철학, 국어국문학, 독어독문학/구조주의 언어학파, 아날학파
2.6. FD_OTHERS	- 기타 학문 분야 및 학파 - 예) 농학, 군사학, 가정학, 건강과학, 박물관학 등
<b>3. THEORY(TR)</b>	- 특정 이론, 법칙, 원리
3.1. TR_SCIENCE	- 과학 관련 이론/법칙/방식/양식/체계/학설 - 예) 21세기형수학공식, 이진법, 원심분리법, 기체확산법
3.2. TR_SOCIAL_SCIENCE	- 사회과학 이론/법칙/방법/원리/사상, 정치사상 - 예) 민주주의, 공산주의, 노동가치설, 대화법, 법치주의, 평등주의
3.3. TR_MEDICINE	- 의학 요법/처방, 의학 진단법 - 예) 인공호흡, 요혈자극법, 바이오피드백
3.4. TR_ART	- 예술 관련 이론/법칙/방식/양식, 예술 사조 - 예) 고딕양식, 고전주의, 몽타주, 오마주, 클로즈업 샷, 루즈샷, 다다이즘
3.5. TR_HUMANITIES	- 철학, 언어, 역사 이론/사상 - 예) 주지주의, 구조주의, 이기론, 관념론, 니힐리즘, 이데아설
3.6. TR_OTHERS	- 기타 이론/법칙/원리 - 예) 샤머니즘, 원칙주의
<b>4. ARTIFACTS(AF)</b>	- 사람에 의해 창조된 인공물
4.1. AF_BUILDING	- 건축물/토목건설물, 운동장, 아파트, 다리, 등대, 분수 등의 명칭 - 예) 대한생명 63빌딩, 동대문운동장, 서해대교, 롯데월드타워
4.2. AF_CULTURAL_ASSET	- 건축물, 고분, 사찰 등의 문화재 명칭 - 예) 송례문, 천마총, 불국사, 서울 원각사지 심층석탑, 예펄탑, 미디 운하
4.3. AF_ROAD	- 도로/철로 명칭 - 예) 세종로, 올림픽대로, 삼성로1길, 충북선, 경부선
4.4. AF_TRANSPORT	- 교통수단/자동차/선박 모델 및 유형, 운송수단, 놀이기구 - 예) 갈라테아호, 경비정, 벤츠, 자기부상열차
4.5. AF_MUSICAL_INSTRUMENT	- 악기 명칭 - 예) 기타, 캐스터네츠, 편경, 팽과리, 대금, 해금
4.6. AF_WEAPON	- 무기 명칭 - 예) 기관단총, 시한폭탄, 곡사포, 고정포, 탄도미사일, 도끼, 칼
4.7. AF_ART_WORKS	- 작품명
4.7.1. AFA_DOCUMENT	- 도서/서적 작품명, 지리서, 지도 - 예) 대동여지도, 동의보감, 한역대장경, 반지의 제왕
4.7.2. AFA_PERFORMANCE	- 춤/무용 작품명, 연극/공연/가극/뮤지컬 제목

	- 예) 농악무, 백조의 호수, 명성황후, 캐츠
4.7.3. AFA_VIDEO	-영화 작품명, TV 프로그램 이름, TV/라디오 채널명 - 예) 기생충, 아빠어디가, CJ홈쇼핑, KBS 2TV
4.7.4. AFA_ART_CRAFT	- 미술 작품명, 조각 등의 조형 작품, 불교 그림명/불상명 - 예) 생각하는 사람, 도산도, 게르니카, 지옥의 문
4.7.5. AFA_MUSIC	- 음악 작품명 - 예) 돈 카를로스, 아이다, 카르멘, 한국환상곡, 보태평
4.8. AF_WARES	- 판매되는 상품 및 제품
4.8.1. AFW_SERVICE_PRODUCTS	- 서비스업의 상품 명칭 - 예) 우리희망드림적금, 에어비앤비, 동유럽 발칸 일주 9일
4.8.2. AFW_OTHER_PRODUCTS	- 기타 공산품의 상품명 - 예) 포스트잇, 에어조던 원, 그린터치 물티슈
<b>5. ORGANIZATION(OG)</b>	- 기관 및 단체
5.1. OGG_ECONOMY	- 경제 관련 기관/단체 - 예) 삼성전자, 현대자동차, LG그룹, OECD, 동양증권
5.2. OGG_EDUCATION	- 교육 기관/단체, 교육 관련 기관 - 예) 밀알유치원, 서울대학교, 청담어학원, 논산훈련소, 국제한국어교육학회
5.3. OGG_MILITARY	- 군사 기관/단체 및 유형, 국방 기관 - 예) 해군, 공군, 육군, 인민군, 해군기지, 국방품질관리소, 한미연합사령부
5.4. OGG_MEDIA	- 미디어 기관/단체, 방송 관련 기관/기업 - 예) KBS, MBC, TJB, 한국일보, 씨네 21, 공익광고협의회, 월스트리트저널, AP통신
5.5. OGG_SPORTS	- 스포츠 기관/단체 - 예) 국민체육진흥심의위원회, 국기원, 국군체육부대, 맨체스터 유나이티드
5.6. OGG_ART	- 예술 기관/단체 - 예) 독일디자인협회, 루브르박물관, 예술의 전당, 리움미술관, 대한극장
5.7. OGG_MEDICINE	- 의학/의료 기관/단체 - 예) 세계보건기구, KMI한국의학연구소, 삼성서울병원, 경희한의원, 새봄약국
5.8. OGG_RELIGION	- 종교 기관/단체, 종파 - 예) 기독교, 불교, 증산도, 대한성공회, 한국기독교장로회, 장로교, 성결교, 화엄종
5.9. OGG_SCIENCE	- 과학 기관/단체 - 예) 국제전기통신연합, 한국표준과학연구원, 한국에너지기술연구원
5.10. OGG_LIBRARY	- 도서관 및 도서관 관련 기관/단체 - 예) 국립중앙도서관, 국회도서관, 바티칸도서관, 법학전문도서관

	관, 일송도서관
5.11. OGG_LAW	- 법률 기관/단체 - 예) 광주지방법원, 대한법무사협회, 저작권심의조정위원회
5.12. OGG_POLITICS	- 정부/행정 기관, 공공기관, 정치 기관 - 예) 청와대, 국회, 대사관, 미래부, 정보통신부, 정부, 우체국, 경찰서, 소방서
5.13. OGG_FOOD	- 음식 관련 업체/회사 - 예) 맥도날드, 김밥천국, 롯데리아, 맘스터치, 스타벅스, 파리바게트
5.14. OGG_HOTEL	- 숙박 관련 업체 - 예) 신라호텔, 한화콘도, 대명리조트, 워커히, 호텔신라, 백담사산장
5.15. OGG_OTHERS	- 기타 기관/단체 - 예) 유색인종향상전국협회, 녹색연합, 참여연대, 블링크, 원잇(ONE IT), 아미
<b>6. LOCATION(LC)</b>	- 지역/장소, 지형/지리
6.1. LCP_COUNTRY	- 국가명 - 예) 한국, 미국, 독일, 프랑스, 터키, 러시아, 가나
6.2. LCP_PROVINCE	- 도, 주 지역명 - 예) 경기도, 제주도, 충청도, 강원도, 전라도, 경상도, 텍사스주
6.3. LCP_COUNTY	- 군, 면, 읍, 리, 동 등과 같은 세부 행정구역명 - 예) 영월군, 도암면, 횡계리, 노암동
6.4. LCP_CITY	- 도시명 - 예) 강릉, 부산, 태백, 뉴욕, 맨체스터, 덴버, 요코하마, 이스탄불
6.5. LCP_CAPITALCITY	- 수도명 - 예) 서울, 베이징, 런던, 파리, 워싱턴 D.C., 오타와, 캔버라
6.6. LCG_RIVER	- 강, 호수, 연못 명칭 - 예) 한강, 섬진강, 춘천호, 아산호, 청계천, 중랑천, 관곡지, 예당저수지
6.7. LCG_OCEAN	- 해양, 바다 명칭 - 예) 남극해, 북극해, 동해, 태평양, 산호해
6.8. LCG_BAY	- 반도, 만 명칭 - 예) 라몬만, 라오동반도, 변산반도, 도쿄만
6.9. LCG_MOUNTAIN	- 산, 산맥, 능선, 고개/재, 봉 명칭 - 예) 라우산, 태백산맥, 차령산맥, 계룡산, 공룡능선, 박달재
6.10. LCG_ISLAND	- 섬, 제도 명칭 - 예) 나이팅게일섬, 로스제도, 백령도, 독도
6.11. LCG_CONTINENT	- 대륙명 - 예) 아시아, 아프리카, 아메리카, 오세아니아, 유라시아, 유럽
6.12. LC_SPACE	- 천체 명칭



	- 예) 태양, 달, 지구, 수성, 모어하우스혜성, 헬리혜성, 전갈자리, 고리성운
6.13. LC_OTHERS	- LC 계열의 세부 유형이 아닌 기타 장소 - 예) 서울역, 신도림역, 인천국제공항, 서울남부터미널, 국립수목원, 금강휴게소
<b>7. CIVILIZATION(CV)</b>	- 문명/문화
7.1. CV_CULTURE	- 문명/문화, 혁명/혁신운동 명칭 - 예) 인더스문명, 황허문명, 거석문화, 도식문화, 4차산업혁명, 르네상스
7.2. CV_TRIBE	- 민족/종족 명칭, 국가를 구성하는 국민 - 예) 이라크족, 유대인, 조선족, 마오리족, 아메리카인디언
7.3. CV_LANGUAGE	- 언어 명칭 - 예) 한국어, 아랍어, 프랑스어, 산스크리트
7.4. CV_POLICY	- 제도/정책 명칭 - 예) 카스트, 쓰레기종량제, 직업공무원제도, 연금제도, 고시제도
7.5. CV_LAW	- 법/법률 명칭 - 예) 국제연합헌장, 민법, 형법, 함무라비법전, 관습법, 헌법
7.6. CV_CURRENCY	- 통화 명칭 - 예) 달러, 원, 엔, 유로, 루피, 위안
7.7. CV_TAX	- 조세 명칭 - 예) 법인세, 소득공제, 면허세, 관세, 부가가치세
7.8. CV_FUNDS	- 연금, 기금, 자금, 펀드 명칭 - 예) 거치연금, 문예진흥기금, 공적자금, 퇴직연금, 코리아 펀드
7.9. CV_ART	- 예술의 분류, 종류 - 예) 소설, 시, 판타지소설, 판소리, 힙합, 케이팝, 공포영화, 미국드라마
7.10. CV_SPORTS	- 스포츠/레포츠/레저 명칭 - 예) 축구, 야구, 수상스키, 스키, 스케이트, 하이킹, 행글라이딩
7.11. CV_SPORTS_POSITION	- 스포츠 포지션 - 예) 투수, 포수, 스트라이커, 골키퍼, 미드필더
7.12. CV_SPORTS_INST	- 스포츠 용품/도구 - 예) 스케이트, 인라인스케이트, 야구공, 배트, 아이젠, 오리발
7.13. CV_PRIZE	- 상, 훈장 명칭 - 예) 노벨경제학상, 대통령상, 대중상, 다승왕, 타격왕
7.14. CV_RELATION	- 가족/친족 관계 명칭 - 예) 엄마, 아버지, 가족, 형제, 제자, 아들, 딸, 자식, 친척
7.15. CV_OCCUPATION	- 직업 명칭 - 예) 화가, 극작가, 건축가, 가정교사, 언어학자, 공인중개사
7.16. CV_POSITION	- 직위/직책 명칭 - 예) 부장, 과장, 대리, 대위, 중위, 소위, 대통령, 국무총리, 장관
7.17. CV_FOOD	- 음식 명칭, 음식 재료

	- 예) 소고기, 소금, 명란젓, 국수, 김치, 나물, 된장찌개, 1955버거, 맵술랭치킨
7.18. CV_DRINK	- 음료, 술 명칭 - 예) 물, 우유, 주스, 홍차, 레모네이드, 칵테일, 위스키, 레드불
7.19. CV_FOOD_STYLE	- 음식 유형 - 예) 한식, 양식, 중식, 보양식, 패스트푸드
7.20. CV_CLOTHING	- 의복/섬유 명칭 - 예) 모시, 메리야스, 와이셔츠, 레이온, 양모, 점퍼
7.21. CV_BUILDING_TYPE	- 건축 양식 명칭 - 예) 기능주의건축, 노르만양식, 시토파건축, 도리스양식
<b>8. DATE(DT)</b>	- 기간 및 계절, 시기/시대
8.1. DT_DURATION	- 기간 - 예) ~부터 ~까지, ~ 간, 전반기/후반기, 성수기/비수기
8.2. DT_DAY	- 날짜/절기 - 예) 입춘, 곡우, 내일, 모레, 금일, 당일, 29일
8.3. DT_WEEK	- 주 - 예) 지난주, 이번 주, 다음 주, 첫 주, 둘째 주
8.4. DT_MONTH	- 달 - 예) 이달, 지난달, 4월, 10월
8.5. DT_YEAR	- 년 - 예) 작년, 내년, 올해, 서기 2910년, 영락1년
8.6. DT_SEASON	- 계절 - 예) 봄, 여름, 가을, 겨울, 춘계, 하계, 추계, 동계
8.7. DT_GEOAGE	- 지질시대 - 예) 원시시대, 구석기시대, 캄브리아기, 중생대, 원생대, 선사시대
8.8. DT_DYNASTY	- 왕조시대 - 예) 청대, 조선시대, 조선 후기, 명대, 고려시대, 고려 말
8.9. DT_OTHERS	- 세부분류 이외의 날짜 - 예) ~년 후/전, ~부터/까지/정도, ~세기, ~때
<b>9. TIME(TI)</b>	- 시계상으로 나타나는 시/시각, 시간 범위
9.1. TI_DURATION	- 기간 - 예) 6시~9시, 낮, 밤, 점심, 저녁, 오전, 오후, 밤낮, 전반전
9.2. TI_HOUR	- 시각 - 예) 12시, 저녁 8시
9.3. TI_MINUTE	- 분 - 예) 13분, 30분, 29분
9.4. TI_SECOND	- 초 - 예) 27초, 15초, 30초
9.5. TI_OTHERS	- 기타 시간

	- 예) 3시 이전, 8시 20분, 6시까지
<b>10. QUANTITY(QT)</b>	- 수량/분량, 순서/순차, 수사로 이루어진 표현
10.1. QT_AGE	- 나이 - 예) 만 55세, 50~60대, 생후 10주, 3개월령
10.2. QT_SIZE	- 넓이, 면적 - 예) 1제곱미터(m²), 1아르(a), 1헥타르(ha), 9, 606.33km²
10.3. QT_LENGTH	- 길이/거리/높이 - 예) 0.76mm, 900m, 8km, 82피트
10.4. QT_COUNT	- 수량/빈도 - 예) 6개, 8백 70선, 하나, 둘
10.5. QT_MAN_COUNT	- 인원수 - 예) 한 명, 두 사람, 3명, 열댓 명
10.6. QT_WEIGHT	- 무게 - 예) 30kg, 140q, 백 근, 9만 t, 85~100kg, 60q, 5만 톤, 170t
10.7. QT_PERCENTAGE	- 백분율, 비율, 농도 - 예) 약 75%, 8~12ppm, 4배, 400~500ppm, 1/3, 절반 이상
10.8. QT_SPEED	- 속도 - 예) 1km/s, 30km/s
10.9. QT_TEMPERATURE	- 온도 - 예) 22~25°C, 41~43.5°
10.10. QT_VOLUME	- 부피 - 예) 두 말, 2홑, 20ℓ, 2되, 2만 2000km³
10.11. QT_ORDER	- 순서, 순차, 등급 - 예) 1위, 3수, 100주년, 제10회, 3학년, 제10회, 제2대
10.12. QT_PRICE	- 금액 - 예) 72만 원, 60억 달러
10.13. QT_PHONE	- 전화번호 - 예) 042-101-1010, 114, 112, 119, 1588-1588, 031, 02, 001, 00700
10.14. QT_SPORTS	- 스포츠 관련 수량 표현 - 예) 1승2패, 5타점, 3:2, 2연승, 6이닝, 1볼넷, 6오버파
10.15. QT_CHANNEL	- TV/라디오 채널 번호 - 예) 2채널, FM 98.1, 91.9, 95.9 MHz
10.16. QT_ALBUM	- 앨범 관련 수량 표현 - 예) 1집, 1.5집, VI, part 1, vol. 2
10.17. QT_ADDRESS	- 주소 관련 숫자 표현 - 예) 110-704, 200-776, 24457, 32호, 179번지, 668-4
10.18. QT_OTHERS	- 기타 수량 표현 - 예) 45점, 100점, 지식머니 20, 북위 38도, 53' 51, 2960X1440
<b>11. EVENT(EV)</b>	- 특정 사건/사고/행사 명칭
11.1. EV_ACTIVITY	- 사회운동 및 선언 명칭

	- 예) 공민권운동, 브나로드운동, 노예해방선언, 인권선언, 밀라노 칙령
11.2. EV_WAR_REVOLUTION	- 전쟁/혁명, 전투 명칭 - 예) 러일전쟁, 남북전쟁, 청교도혁명, 4.19혁명, 명량해전, 516 쿠데타
11.3. EV_SPORTS	- 스포츠/레저 관련 행사 명칭 - 예) 올림픽경기대회, 서울올림픽, 오버워치리그, 태권도 대회
11.4. EV_FESTIVAL	- 축제, 영화제, 콘서트, 행사 명칭 - 예) 이천도자기축제, 칸 영화제, 방탄 콘서트, 책 읽는 춘천, 서울 패션 위크
11.5. EV_OTHERS	- 사건/사고/사태, 회의/학술대회, 전시회, 경진대회 명칭 - 예) 세월호 참사, 다보스포럼, 키스해링전, 월드바리스타챔피언쉽
<b>12. ANIMAL(AM)</b>	- 사람을 제외한 짐승
12.1. AM_INSECT	- 곤충 명칭 - 예) 버메뚜기, 가시거미, 지네, 불나방, 배추흰나비, 참매미
12.2. AM_BIRD	- 조류 명칭 - 예) 왜가리, 뺨꾸기, 거위, 두루미, 황새
12.3. AM_FISH	- 어류 명칭 - 예) 참가자미, 전기뱀장어, 연어, 갈치, 멸치, 미꾸라지, 송어, 날치
12.4. AM_MAMMALIA	- 포유류 명칭 - 예) 해달, 토끼, 오소리, 돌고래, 박쥐, 고슴도치, 두더쥐
12.5. AM_AMPHIBIA	- 양서류 명칭 - 예) 개구리, 두꺼비, 산개구리, 산청개구리, 도롱뇽, 맹꽁이
12.6. AM_REPTILIA	- 파충류 명칭 - 예) 도마뱀, 거북, 자라, 구렁이, 백사, 공룡, 악어
12.7. AM_TYPE	- 동물 분류 명칭 - 예) 강장동물, 구두동물, 극피동물, 내향동물, 모약동물, 선형동물
12.8. AM_PART	- 동물이나 사람 몸의 한 부분(신체 부위) 명칭 - 예) 목, 살, 등, 배, 몸통, 손, 입, 갈기
12.9. AM_OTHERS	- 기타 동물 명칭 - 예) 가재(절지동물), 문어, 오징어(연체동물)
<b>13. PLANT(PT)</b>	- 꽃/나무, 육지식물, 해조류, 버섯류, 이끼류
13.1. PT_FRUIT	- 과일이나 열매 이름 - 예) 사과, 배, 자두, 바나나, 복숭아, 오렌지, 포도, 호두, 쌀, 보리
13.2. PT_FLOWER	- 꽃 이름 - 예) 저면아이리스, 백합, 국화, 제비꽃, 히아신스, 초롱꽃, 할미꽃

13.3. PT_TREE	- 나무 이름 - 예) 후박나무, 벚나무, 분비나무, 너도밤나무
13.4. PT_GRASS	- 풀 이름 - 예) 향부자, 하늘타리뿌리, 인삼, 생강, 감자, 싱아, 갈대
13.5. PT_TYPE	- 식물 분류 명칭 - 예) 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
13.6. PT_PART	- 식물 각 부분의 명칭 - 예) 뿌리, 가지, 가시, 꽃, 꽃잎, 꽃봉오리, 꽃수술, 꽃받침, 씨앗
13.7. PT_OTHERS	- 기타 식물 명칭 - 예) 우산이끼, 송이버섯, 김, 미역, 매생이, 선인장, 콩나물
<b>14. MATERIAL(MT)</b>	- 원소 및 금속, 암석/보석, 화학물질
14.1. MT_ELEMENT	- 원소명 - 예) 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨
14.2. MT_METAL	- 금속 명칭 - 예) 황산철, 알루미늄, 아연, 구리, 수은, 아말감
14.3. MT_ROCK	- 암석 명칭 - 예) 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암
14.4. MT_CHEMICAL	- 화학 물질 명칭 - 예) 암모니아, 과산화수소, 에탄올, 황화수소, 네온, 일산화탄소, 헤모글로빈, 비타민
<b>15. TERM(TM)</b>	- 위에서 정의된 세부 개체명 이외의 개체명
15.1. TM_COLOR	- 색깔 - 예) 흰색, 흰빛, 흑색, 회색, 황, 홍색, 하늘색, 파란빛
15.2. TM_DIRECTION	- 방향 - 예) 최북단, 중앙부, 중앙, 중부, 서쪽, 서부, 상부, 북쪽
15.3. TM_CLIMATE	- 기후 지역 명칭 - 예) 아열대기후, 고산기후, 고지기후, 도시기후, 동안기후
15.4. TM_SHAPE	- 모양/형태/무늬 명칭 - 예) 삼각형, 사각형, 팔각형, 일자형, 원통형, 삼각뿔 모양, 가로 무늬, 구름무늬
15.5. TM_CELL_TISSUE_ORGAN	- 세포/조직/기관 명칭(외부에서 보이지 않는 조직이나 기관) - 예) 간상세포, 각막세포, 고막, 뉴런, 늑연골, 식도, 콩팥, 비장
15.6. TMM_DISEASE	- 증상/증세/질병 명칭 - 예) 기침, 폐결핵, 기관지확장증, 레트증후군, 피고름가래
15.7. TMM_DRUG	- 약/약품명 - 예) 헤파린, 근이완제, 모기향, 소화제, 아스피린, 구충제
15.8. TMI_HW	- IT 하드웨어 명칭, 장치/부품명 - 예) 하드디스크드라이브, 마이크로프로세서, 롬, 디스켓, 소자 소켓, 감각 센서
15.9. TMI_SW	- IT 소프트웨어 용어

	- 예) 윈도우10, 리눅스, 반디집, 알송, 한컴오피스 한글, MS워드, 트로이목마, 자바
15.10. TMI_SITE	- URL 주소 - 예) <a href="https://opendict.korean.go.kr/">https://opendict.korean.go.kr/</a> , <a href="https://www.korean.go.kr/">https://www.korean.go.kr/</a>
15.11. TMI_EMAIL	- 이메일 주소 - 예) esuhyeon@glab.hallym.ac.kr
15.12. TMI_MODEL	- 각종 제품의 세부 모델명, 모델일련번호 - 예) LM1000-2CX
15.13. TMI_SERVICE	- IT 서비스 용어 - 예) 와이브로서비스, DMB서비스, 카카오톡, 인스타그램, 네이버, 다음, 구글
15.14. TMI_PROJECT	- 프로젝트 명칭 - 예) BrainKorea21, 4대강사업, 그린 IT 사업, 무료치과진료사업
15.15. TMIG_GENRE	- 게임 장르 명칭 - 예) RPG, 실시간 전략 게임, 1인칭 슈팅 게임
15.16. TM_SPORTS	- 스포츠/레저 용어(기술/규칙 명칭) - 예) 밀어내기 득점, 솔로 홈런, 바스켓카운트, 리바운드, 발리슛

이 지침에서는 위의 대분류 15 항목과 세부분류 150 항목의 태그로 개체명을 태깅하여 개체명 분석 말뭉치를 구축하는 방안을 제시할 것이다. 개체명 말뭉치를 구축할 때 주의해야 할 사항은 세 가지가 있다. 첫째는 개체명 태깅 대상, 둘째는 개체명 태깅 단위, 셋째는 개체명 태깅 범주이다.

첫 번째는 '개체명은 무엇을 대상으로 하는가', 즉 개체명 태깅 대상에 관한 문제이다. 개체명(Named-Entity)은 다른 것들과 구별하여 부르기 위해 이름을 붙인 실제 세계에 존재하는 개체들의 명칭이다. 따라서 개체명에는 대부분의 고유명이 포함된다. 그러나 개체명 태깅 대상을 고유명으로 한정하면 인명, 지명, 작품명, 단체명 등 특정 범주에 속하는 개체명들로 그 대상이 제한된다. 따라서 인명, 지명, 작품명 등 이외의 범주는 고유명이 아닌 세부분류명(보통 명사)이 개체명 태깅 대상에 포함된다.

두 번째는 '어디부터 어디까지 태깅할 것인가', 즉 개체명 태깅 단위에 관한 문제이다. 단어 뿐 아니라 '구'도 개체명이 될 수 있으므로, 단어 단위로 태깅할 것인지 명사구 전체를 하나로 태깅할 것인지가 문제가 된다. 일례로 '한국해양대학'과 같이 국가명이 고유명에 포함되어 있는 경우, '한국'과 '해양대학'을 따로 태깅하면 본래의 의미를 갖지 못하므로 <한국해양대학: OGG\_EDUCATION>과 같이 전체를 하나로 태깅한다. 그러나 '대전 유성구 신성동'의 경우, <대전 유성구 신성동: LC\_OTHERS>과 같이 태깅할 수도 있고 <대전: LCP\_CITY> <유성구: LCP\_COUNTY> <신성동: LCP\_COUNTY>과 같이 태깅할 수도 있다. 이 경우 명사구 전체의 의미가 각 단어의 의미의 합과 같다면, 즉 구의 의미가 투명하다면 최소 단위로 태깅하는 방식을 취한다. 따라서 '대전 유성구 신성동'은 두 번째 방식으로 태깅한다.

개체명의 태깅 단위에서 한 가지 더 주의할 사항은 복합어(합성어, 파생어) 내부의 구성요소는 개체명에 해당되더라도 따로 태깅하지 않는다는 점이다. 복합어는 그 자체로 해당 복합어의 구성요소와는 구별되는 별도의 개체를 지시하므로, 복합어 전체가 개체명 태깅 대상일 경

우에만 태깅하며, 복합어가 개체명 태깅 대상이 아닐 때에는 내부 구성요소가 개체명 태깅 대상이더라도 별도로 태깅하지 않는다. 예를 들어 '야구장'의 '야구', '경찰차'의 '경찰'은 둘 다 CV에 속하는 태깅 대상이나, 합성어 내부의 요소이므로 태깅하지 않는다. 일부 예외가 되는 사례가 있는데 이는 뒤에서 자세히 살필 것이다.

세 번째는 '개체명에 어떤 태그를 부여할 것인가', 즉 태깅 범주에 관한 문제이다. 어떤 단어 혹은 구가 개체명으로 판정되었다면, 그것의 의미 범주에 부합하는 태그를 달아 주어야 한다. 그러나 일부 경우 하나의 개체가 둘 이상의 의미 범주에 속하는 경우가 있다. 예를 들어 '시금치'의 경우 '식물'이면서 '음식물'이므로 PT\_GRASS와 CV\_FOOD에 모두 해당된다. 그러나 유일 태깅을 원칙으로 하므로 문맥과 적정술어를 고려하여 가장 적절한 한 가지로 태깅을 해야 한다. 간혹 문맥으로 판정하는 것이 어려운 경우도 있는데, 이러한 유형은 따로 목록화하여 우선순위를 배정할 것이다.

이 지침에서는 각 개체명 부류에 대하여 태깅 대상, 태깅 단위, 태깅 범주를 구체적으로 제시할 것이다.

## 나. 태깅 대상

---

여기에서는 '개체명 태깅은 무엇을 대상으로 하는가', 즉 개체명 태깅 대상을 상론한다.

- 개체명의 태깅 대상은 기본적으로 고유명이다. 사전적으로 고유명은 '날날의 특정한 사물이나 사람을 다른 것들과 구별하여 부르기 위해 붙인 고유의 이름'을 의미하는 것으로, 실제 세계에 유일하게 존재하는 특정 '실체'의 명칭을 나타낸다. 그러나 개체명 태깅 대상을 고유명으로 한정하면 인명, 지명, 작품명, 단체명 등 특정 범주에 속하는 개체명들로 그 대상이 제한된다. 또한 자연종명 등 일부 부류는 보통명사인지 고유명인지에 대한 구분이 명확하지 않은 측면이 있다. 따라서 인명, 지명, 작품명 등 이외의 범주는 고유명이 아닌 세부분류명이 개체명 태깅 대상에 포함된다. 예를 들어 AM과 PT에는 동물과 식물의 다양한 종이 개체명 대상이 되며, MT에는 원소나 금속, 암석의 구체적인 유형이, FD에는 학문의 하위 분야가 개체명의 대상이 되는 등이다. 이 사업에서 '개체명 태깅 대상'에 해당하는 항목들을 정리하면 다음과 같다.

- 고유명

- 인명
- 건축물명
- 도로명
- 작품명
- 상품명
- 기관명
- 단체명
- 국가명
- 상/훈장명
- 지명
- 사회운동/선언의 명칭
- 전쟁/혁명/전투의 명칭
- 스포츠 관련 행사의 명칭
- 축제/영화제/콘서트의 명칭
- 전쟁/혁명의 명칭
- 제품의 모델명
- IT 서비스 명칭
- IT/국가/기업 프로젝트 명칭

- 보통명사

- 특정 대상을 지시하는 보통명사



- 문명/문화
- 제도/정책
- 이론/법칙/원리
- 통화
- 인간관계
- 악기
- 스포츠 도구/용품
- 날짜/시간/기간
- 순서/수량
- 동물
- 식물
- 신체 부위
- 식물의 조직/부분
- 원소/금속/암석
- 색깔
- 방향
- 세포/조직/기관
- 증상/질환
- 약/의약품
- 스포츠 기술/규칙
- 유형을 나타내는 보통명사
  - 유형을 나타내는 보통명사가 태깅 대상인 대분류
    - STUDY\_FIELD
    - THEORY
  - 유형을 나타내는 보통명사가 태깅 대상인 세분류
    - AF\_WEAPON (칼, 도끼 등)
    - AF\_TRANSPORT (자동차, 비행기 등)
    - OGG\_MILITARY (해군, 공군, 육군 등)
    - OGG\_RELIGION (도교, 불교, 가톨릭 등)
    - OGG\_POLITICS (우체국, 경찰 등)
    - CV\_TRIBE (집시, 백인, 유대인 등)
    - CV\_SPORTS (스케이팅, 스키, 축구 등)
    - CV\_SPORTS\_POSITION (가드, 포워드, 수비수, 골키퍼 등)
    - CV\_ART (공포영화, 미국 드라마 등)
    - CV\_BUILDING\_TYPE (노르만 양식, 기와집, 한옥 등)
    - CV\_FOOD (국, 탕국, 밥, 무침, 향신료 등)

- CV\_DRINK (우유, 주스, 술 등)
- CV\_CLOTHING (티셔츠, 점퍼, 와이셔츠 등)
- CV\_POSITION (부장, 과장, 대통령 등)
- CV\_OCCUPATION (화가, 건축가, 언어학자, 아나운서 등)
- CV\_LAW (헌법, 선거법, 폭행죄, 명예훼손죄 등)
- CV\_TAX (근로소득세, 부가가치세, 취득세, 재산세 등)
- CV\_FUNDS (개인연금, 노령연금, 통화안정기금, 방송발전기금 등)
- CV\_FOOD\_STYLE (한식, 일식, 양식, 중식, 보양식, 채식 등)
- EV\_SPORTS (마라톤대회, 등반대회, 축구대회 등)
- EV\_FESTIVAL (장미축제, 연등축제, 단풍축제, 록페스티벌 등)
- AM\_TYPE (강장동물, 절지동물, 해면동물, 포유류, 어류 등)
- PT\_TYPE (녹조류, 장미목, 붓꽃과, 낙엽교목, 양치식물 등)
- TM\_CLIMATE (열대기후, 건조기후, 온대기후 등)
- TM\_SHAPE (유선형, 일자형, 원뿔형, 삼각형, 사각형, 가로무늬, 줄무늬 등)
- TMM\_DRUG (영양제, 각성제, 건강제, 수면제, 연고 등)
- TMI\_HW (디지털카메라, 노트북, 전자레인지 등)
- TMI\_SW (윈도우, 리눅스, 익스플로러, 크롬 등)
- TMIG\_GENRE (RPG, 실시간 전략 게임, 1인칭 슈팅 게임, 모바일 게임 등)

## 다. 태깅 단위

여기에서는 개체명을 태깅할 때 주의해야 할 사항인 개체명 태깅 단위에 대한 원칙을 상론한다. 앞에서 개체명의 태깅 대상에 고유명뿐만 아니라 보통명사도 포함된다고 하였다. 고유명과 보통명사는 속성이 다르므로, 태깅 단위 결정의 원리는 동일하더라도 그 원리를 적용함에 있어서는 부분적으로 차이가 있을 수밖에 없다. 따라서 여기에서는 태깅 단위 결정의 원리를 고유명과 보통명사로 나누어 제시할 것이다.

### 1. 태깅 단위 결정의 원리 - 고유명

- 개체명은 명사 또는 명사구를 대상으로 하므로 명사구 전체를 하나로 태깅할 것인지, 단어 단위로 나누어서 태깅할 것인지 문제가 되는 경우가 많다. 명사구로 구성된 고유명의 경우, 명사구에 포함된 각각의 단어들이 해당 개체를 지시하는 것이 아니라 명사구 전체가 하나로 해당 개체를 지시하는 것이기 때문에 원칙적으로 '최장 단위 태깅'을 한다.
- 실용적인 목적을 위해 내부 구성요소의 정보를 참조할 필요가 있는 경우, 전체 명사구의 의미가 합성적이어서 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되는 경우에만 예외적으로 '최소 단위 태깅'을 한다.

#### 1.1. 최장 단위 태깅(원칙)

- 개체명 태깅의 대상이 되는 고유명은 단어이든 구이든 그 전체가 하나의 대상을 지시하므로 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
- 예를 들어, '경찰청 야구단'의 경우, '경찰청'과 '야구단' 각각의 기관을 의미하는 것이 아니라, 전체가 특정 스포츠 팀의 명칭이다. 또한 각각을 따로 태깅하게 되면 '야구단'은 태깅 대상이 아니어서 '경찰청'만 태깅하게 되므로 '경찰청 야구단'의 본유적 의미인 '스포츠 팀'에 해당하는 태깅을 할 수 없게 된다. 따라서 이러한 경우에는 '경찰청 야구단' 전체를 OGG\_SPORTS로 태깅한다.

##### ○ 예제

- 한국해양대학
  - '한국해양대학'의 경우, 대학 이름에 '한국'이라는 지명이 포함된 것이므로 함께 태깅하지 않으면 본래 단체명이 훼손된다. 따라서 전체를 하나로 태깅한다.
  - <한국해양대학: OGG\_EDUCATION> (O)
  - <한국: LCP\_COUNTRY> <해양 대학: OGG\_EDUCATION> (X)
- 스페인 대사관
  - '스페인 대사관'은 대사관이 소재하고 있는 위치를 의미하는 것이 아니므로 통합하여 태깅한다.

- <스페인 대사관: OGG\_POLITICS> (O)
  - <스페인: LCP\_COUNTRY> <대사관: OGG\_POLITICS> (X)
- 콩순이 말하는 냉장고
  - '콩순이 말하는 냉장고'와 같은 상품명인 경우도 제품 이름에 '콩순이'라는 캐릭터명이 포함된 것이므로, 따로 태깅하지 않고 전체를 하나로 태깅한다.
  - <콩순이 말하는 냉장고: AFW\_OTHER\_PRODUCTS> (O)
  - <콩순이: PS\_CHARACTER> <말하는 냉장고: AFW\_OTHER\_PRODUCTS> (X)
- 제24회 서울 올림픽 경기
  - 스포츠 행사는 보통 개최 횟수와 개최지가 함께 출현하는데, 개최 횟수(QT)는 구분하여 태깅하나 행사가 열리는 지역(LC)은 통합하여 태깅한다.
  - <제24회: QT\_ORDER> <서울 올림픽 경기: EV\_SPORTS> (O)
  - <제24회: QT\_ORDER> <서울: LCP\_CAPITALCITY> <올림픽 경기: EV\_SPORTS> (X)
  - <제24회 서울 올림픽 경기: EV\_SPORTS> (X)
- 88 서울 올림픽 경기
  - '1988년'을 줄여서 '88'과 같이 표현한 경우 '88'만으로는 '1988년'임을 보여줄 수 없다. 따라서 개최 연도의 일부만 쓰인 경우는 후행하는 EV와 함께 태깅한다.
  - <88 서울 올림픽 경기: EV\_SPORTS> (O)
  - <88: DT\_YEAR> <서울 올림픽 경기: EV\_SPORTS> (X)
- 단어 단위로 나누어 태깅하면 특정 개체를 지칭하지 못하게 되는 경우, 즉 다수의 개체를 지칭하게 되는 경우 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
  - 예제
    - 현대오일뱅크 대산공장
      - '현대오일뱅크 대산공장'의 경우, '현대오일뱅크'와 '대산공장'을 따로 태깅하면 '대산공장'이 '현대오일뱅크'의 '대산공장'인지, '롯데케미칼'의 '대산공장'인지 불분명해지므로 전체를 하나로 태깅한다.
      - <현대오일뱅크 대산공장: LC\_OTHERS> (O)
      - <현대오일뱅크: OGG\_ECONOMY> <대산공장: LC\_OTHERS> (X)
    - 스타벅스 대학로점
      - '스타벅스 대학로점'도 마찬가지로 '대학로점'만으로는 어떤 기관의 '대학로점'을 지칭하는지 알 수 없으므로, 전체를 하나로 태깅한다.
      - <스타벅스 대학로점: OGG\_FOOD> (O)
      - <스타벅스: OGG\_FOOD> <대학로점: OGG\_FOOD> (X)

## 1.2. 최소 단위 태깅(예외)

- 실용적인 목적을 위해 내부 구성 요소의 정보를 참조할 필요가 있는 경우, '최소 단위 태

킹'을 하여 각 단어를 따로 태깅한다. 지역명(LC)의 연속 혹은 지역명과 그곳에 위치한 건축물(AF)이나 그 지역을 본거지로 하는 조직(OG)은 각기 따로 태깅한다. 특정 행사(EV)나 상(CV)이 날짜(DT)나 횟수(QT)와 함께 쓰일 때에도 구분하여 태깅한다.

○ 예제

■ LC + LC

■ 대전 유성구 신성동

■ <대전: LCP\_CITY> <유성구: LCP\_COUNTY> <신성동: LCP\_COUNTY> (O)

■ <대전 유성구 신성동: LC\_OTHERS> (X)

■ LC + AF

■ 잠실 롯데월드타워

■ <잠실: LCP\_COUNTY> <롯데월드타워: AF\_BUILDING> (O)

■ <잠실 롯데월드타워: AF\_BUILDING> (X)

■ LC + OGG

■ 미국 컬럼비아대학

■ <미국: LCP\_COUNTRY> <컬럼비아대학: OGG\_EDUCATION> (O)

■ <미국 컬럼비아대학: OGG\_EDUCATION> (X)

■ 대한제국 법부 전라북도재판소

■ <대한제국: LCP\_COUNTRY> <법부: OGG\_LAW> <전라북도재판소: OGG\_LAW> (O)

■ <대한제국 법부 전라북도재판소: OGG\_LAW> (X)

■ DT/QT + CV/EV

■ 2000년 제34회 경제정의기업상

■ <2000년: DT\_YEAR> <제34회: QT\_ORDER> <경제정의기업상: CV\_PRIZE> (O)

■ <2000년: DT\_YEAR> <제34회 경제정의기업상: CV\_PRIZE> (X)

■ 제24회 서울올림픽경기

■ <제24회: QT\_ORDER> <서울올림픽경기: EV\_SPORTS> (O)

■ <제24회 서울올림픽경기: EV\_SPORTS> (X)

■ PS + CV

■ 문재인 대통령

■ <문재인: PS\_NAME> <대통령: CV\_POSITION> (O)

■ <문재인 대통령: CV\_POSITION> (X)

- 해당 명사구가 정식 명칭이 아닌 경우, 전체를 하나로 태깅하지 않고 '최소 단위 태깅'을 하여 각 단어를 따로 태깅한다.

○ 예제

■ 광주 돔구장

- '광주 돔구장'은 정식 명칭이 아니므로 각각 따로 태깅하되, '돔구장'은 개체명이 아니므로 태깅하지 않고 '광주'만 태깅한다.
- <광주: LCP\_CITY> 돔구장
- 서울 올림픽공원 체조경기장
  - '서울 올림픽공원 체조경기장'은 정식 명칭이 아니므로 각각 따로 태깅하되, '체조경기장'은 개체명이 아니므로 태깅하지 않는다.
  - <서울: LCP\_CAPITALCITY> <올림픽공원: LC\_OTHERS> 체조경기장

## 2. 태깅 단위 결정의 원리 – 보통명사

개체명 태깅 대상에는 고유명뿐만 아니라 보통명사도 포함되는데 개체명의 태깅 대상이 되는 보통명사 또한 한 단어나 여러 개의 단어로 이루어져 있다. 보통명사의 경우 사전에 한 단어로 등재되어 있는지 여부가 태깅 단위를 결정함에 있어 중요하게 작용한다. <우리말샘> 사전 검색 결과 해당 명사 구성이 '한 단어'로서 등재되어 있는 경우('-', '^'와 같은 기호와 함께 등재되어 있는 경우) 전체를 하나로 태깅한다. <우리말샘>에 '구'로서 등재되어 있거나(띄어쓰기가 되어 있는 경우), 등재되어 있지 않은 경우에는 세부 지침에 따라 전체를 하나로 태깅하거나, 각각을 따로 태깅한다. 이때 '단어'보다 작은 단위인 복합어(합성어, 파생어) 내부의 구성 요소는 개체명에 해당되더라도 따로 태깅하지 않는다.

### 2.1. 최장 단위 태깅(원칙)

- 특정 학문 분야나 이론/법칙과 같은 전문용어의 경우 그 전체가 하나의 대상을 지시하므로 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
  - 예제
    - 신고전학과 성장론
      - '신고전학과 성장론'은 그 전체가 이론의 명칭이다. 이 명사구 내부에는 '신고전학과'라는 FD\_SOCIAL\_SCIENCE가 포함되어 있으나 '성장론'만으로는 특정한 이론을 지시할 수 없게 되므로 이러한 경우에는 '신고전학과 성장론' 전체를 TR\_SOCIAL\_SCIENCE로 태깅한다.
      - <신고전학과 성장론: TR\_SOCIAL\_SCIENCE> (O)
      - <신고전학과: FD\_SOCIAL\_SCIENCE> 성장론 (X)
- 예술의 분류나 종류, 여러 장르가 혼합되어 새로운 장르가 만들어진 경우 그 전체가 하나의 예술 분류를 지시하므로 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
  - 예제
    - 어쿠스틱 발라드
      - '어쿠스틱'과 '발라드'를 각각 따로 태깅하면 발라드의 한 종류인 '어쿠스틱 발라드'를 지시하지 못하므로 '최장 단위 태깅 원칙'을 적용하여 전체를 하나로 태깅

한다.

- <어쿠스틱 발라드: CV\_ART> (O)
- 어쿠스틱 <발라드: CV\_ART> (X)
- 스포츠나 레저의 명칭 및 종목, 여러 종목이 혼합되어 새로운 종목이 만들어진 경우 그 전체가 하나의 새로운 종목을 지시하므로 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
  - 예제
    - 오픈워터 수영
      - '오픈워터 수영'에서 '수영'만을 태깅하게 되면, 수영의 새로운 하위 종목을 지시하지 못하므로 통합하여 태깅한다.
      - <오픈워터 수영: CV\_SPORTS> (O)
      - 오픈워터 <수영: CV\_SPORTS> (X)
- 친족 관계를 나타내는 명칭의 경우, 전체를 통합해서 태깅해야만 해당 구 구성이 의미하는 대상을 지시할 수 있으므로, 구 단위로 사전에 등재되어 있더라도 각각을 따로 태깅하지 않고 '최장 단위 태깅 원칙'을 적용하여 전체를 하나로 태깅한다.
  - 예제
    - 사촌 동생
      - <사촌 동생: CV\_RELATION> (O)
      - <사촌: CV\_RELATION> <동생: CV\_RELATION> (X)
    - 사촌 누나
      - <사촌 누나: CV\_RELATION> (O)
      - <사촌: CV\_RELATION> <누나: CV\_RELATION> (X)
- 사람이나 동물이 먹는 음식, 간식 등은 각각 따로 태깅하면 명사구 전체의 의미를 드러내지 못하므로 '최장 단위 태깅 원칙'을 적용하여 태깅한다.
  - 예제
    - 송아지 목뼈 껌
      - '송아지', '목뼈', '껌'을 각각 따로 태깅하면 '송아지의 목뼈로 만든 반려동물용 껌'의 의미를 나타내지 못하므로 전체를 하나로 태깅한다.
      - <송아지 목뼈껌: CV\_FOOD> (O)
      - <송아지: AM\_MAMMALIA> <목뼈: TM\_CELL\_TISSUE\_ORGAN> 껌 (X)
    - 봄배추 김치
      - <봄배추 김치: CV\_FOOD> (O)
      - <봄배추: PT\_GRASS> <김치: CV\_FOOD> (X)
- 의복의 유형이나 명칭 또한 따로 태깅하면 명사구 전체의 의미를 드러내지 못하므로 '최장 단위 태깅 원칙'을 적용하여 태깅한다.
  - 예제
    - 카무플라주 점퍼

- '카무플라주'와 '점퍼'를 따로 태깅하면 '카무플라주 무늬의 점퍼'의 의미를 드러 내지 못하게 되므로 전체를 하나로 태깅한다.
  - <카무플라주 점퍼: CV\_CLOTHING> (O)
  - <카무플라주: TM\_SHAPE> <점퍼: CV\_CLOTHING> (X)
- '년/월/일/시, 년/월, 월/일'과 같은 날짜나 시간 표현의 경우(요일 제외)도 단어 단위로 태깅하면 특정 시간이나 특정 날짜를 지칭하지 못하므로 전체를 DT\_OTHERS로 태깅한다.
  - 예제
    - 2019년 10월 29일
      - <2019년 10월 29일: DT\_OTHERS> (O)
      - <2019년: DT\_YEAR> <10월: DT\_MONTH> <29일: DT\_DAY> (X)
- 날짜나 시간 등의 순서 표현도 통합하여 하나로 태깅한다.
  - 예제
    - 이번 주
      - <이번 주: DT\_WEEK> (O)
    - 둘째 주
      - <둘째 주: DT\_WEEK> (O)

## 2.2. 최소 단위 태깅(예외)

- 실용적인 목적을 위해 내부 구성요소의 정보를 참조할 필요가 있는 경우, 전체 명사구의 의미가 합성적이어서 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되는 경우에는 예외적으로 '최소 단위 태깅'을 한다.
- 해당 명사 구성이 구 단위로 <우리말샘> 사전에 등재되어 있거나(띄어쓰기가 되어 있는 경우) 등재되어 있지 않은 경우, 내부 요소를 각각 따로 태깅한다.
  - 예제
    - 힙합 댄서
      - <힙합: CV\_ART> <댄서: CV\_OCCUPATION> (O)
      - <힙합 댄서: CV\_OCCUPATION> (X)
  - 예제
    - 상송 가수
      - <상송: CV\_ART> <가수: CV\_OCCUPATION> (O)
      - <상송 가수: CV\_OCCUPATION> (X)
  - 예제
    - 커피 자판기
      - <커피: CV\_DRINK> <자판기: TMI\_HW> (O)
      - <커피 자판기: TMI\_HW> (X)
- '양치식물'과 같은 유형, 종명과 '고사리'와 같은 해당 유형에 속하는 개체가 함께 쓰인 경



우, 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되므로 최소 단위 태깅을 하여 각각 따로 태깅한다.

○ 예제

■ 양치식물 고사리

■ <양치식물: PT\_TYPE> <고사리: PT\_GRASS> (O)

■ <양치식물 고사리: PT\_GRASS> (X)

- 사람 혹은 동물과 그 대상의 신체 부위 명사가 함께 쓰인 경우, 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되므로 각각 따로 태깅한다.

○ 예제

■ 엄마 손

■ <엄마: CV\_RELATION> <손: AM\_PART> (O)

■ <엄마 손: AM\_PART> (X)

■ 호랑이 이빨

■ <호랑이: AM\_MAMMALIA> <이빨: AM\_PART> (O)

■ <호랑이 이빨: AM\_PART> (X)

- 'N주+요일', '날짜+요일' 구성은 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되므로 각각 따로 태깅한다.

○ 예제

■ 12월 4일 금요일

■ <12월 4일: DT\_OTHERS> <금요일: DT\_DAY> (O)

■ <12월 4일 금요일: DT\_OTHERS> (X)

■ 셋째 주 수요일

■ <셋째 주: DT\_WEEK> <수요일: DT\_DAY> (O)

■ <셋째 주 수요일: DT\_OTHERS> (X)

- '6학년 1반'과 같은 경우, '6학년'과 '1반'이 각각 순서를 나타내며, 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되므로 각각 따로 태깅한다.

○ 예제

■ <6학년: QT\_ORDER> <1반: QT\_ORDER> (O)

■ <6학년 1반: QT\_ORDER> (X)

## 2.3. 주의 사항

- 둘 이상의 단어가 만나 합성어를 이루거나 단어와 접사가 만나 파생어를 이룬 경우, 그 전체가 하나의 개념을 지시하므로 복합어 전체가 개체명일 경우에만 태깅하고 복합어 내부 요소는 개체명이더라도 별도로 태깅하지 않는다.

○ 예제

■ 손바닥: 복합어가 개체명이므로 전체를 AM\_PART로 태깅한다.

- 손짓: 복합어가 개체명이 아니므로 '손짓'은 태깅하지 않는다. 내부 요소 '손'도 태깅하지 않는다.
- 야구장: 복합어가 특정한 인공물을 지시하지 않아 태깅하지 않고, 내부 요소인 '야구'도 태깅하지 않는다.
- 한국적: '한국적 문화'와 같은 표현에서 '한국적'은 개체명이 아니므로 태깅하지 않고, 내부 요소 '한국' 역시 태깅하지 않는다.
- 국어국문학과: '국어국문학과'는 어떤 대학교의 하위 부서이므로 태깅하지 않고, 내부 요소인 '국어국문학' 역시 태깅하지 않는다.
- 다음의 경우는 예외적으로 복합어 내부 요소를 태깅하므로 주의를 요한다.
  - 원소명, 금속물 등 MT는 해당 원소명 및 금속물 등의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다.
    - 예) 마그네슘리본 → <마그네슘: MT\_METAL>리본
  - 다음의 접미사가 결합한 경우, 예외적으로 내부 요소를 태깅한다.
    - 첫째, 접미사 '-들'은 선행 요소의 의미에 영향을 주지 않으므로 '-들'이 결합한 단어는 '-들'을 제외하고 태깅한다.
      - 예) 학생들 → <학생: CV\_OCCUPATION>들
    - 둘째, 접미사 '-님'이 결합한 단어는 전체가 <우리말샘>에 등재된 경우(즉, 관습적으로 '-님'을 붙여서 직위·직책명으로 사용하는 경우) 전체를 태깅하나 그렇지 않은 경우는 '-님'을 제외하고 태깅한다.
      - 예) 선생님 → <선생님: CV\_OCCUPATION>, 회장님 → <회장: CV\_POSITION>님
- 원문의 띄어쓰기 여부와 관계없이 사전(우리말샘 기준)에 한 단어로 등재되어 있으면 전체를 통합하여 태깅한다. 우리말샘 표제항에 '-', '^'와 같은 기호와 함께 등재되어 있는 경우 복합어(준복합어)인 것으로 보고 전체를 통합하여 태깅하고, 띄어쓰기가 되어 있거나 등재되어 있지 않은 경우에는 지침에 따라 전체를 하나로 태깅하거나, 각각을 따로 태깅한다.
  - 예제
    - 소 간, 손가락 뼈: 원문의 띄어쓰기와 관계없이 '소간, 손가락뼈'가 모두 사전에 등재되어 있으므로 통합하여 태깅한다.
    - 소 간 → <소 간: TM\_CELL\_TISSUE\_ORGAN>
    - 손가락 뼈 → <손가락 뼈: TM\_CELL\_TISSUE\_ORGAN>
- 사전(우리말샘)에 구 단위로 등재되어 있지만 기호 '-'와 함께 등재되어 있는 경우, 복합어에 준하는 것으로 보아 내부 태깅을 하지 않고 통합하여 태깅한다.
  - 예제
    - 정부조직법, 법원조직법: 우리말샘 사전에 '정부^조직법', '법원^조직법'으로 등재되어 있으므로 내부 태깅을 하지 않고 통합하여 CV\_LAW로 태깅한다.
      - 정부조직법 → <정부조직법: CV\_LAW>
      - 법원조직법 → <법원조직법: CV\_LAW>

- 사전(우리말샘)에 구 단위로 등재되어 있는 경우, 즉 띄어쓰기가 되어 있는 경우는 한 단어(복합어)가 아니므로, 명사구 전체는 개체명이 아니어서 태깅하지 않더라도 내부 요소는 별도로 태깅한다.
  - 예제
    - 고양이 카페: '고양이 카페'는 우리말샘 사전에 구 단위인 '고양이 카페'로 등재되어 있다(띄어쓰기가 되어 있음). 따라서 개체명인 '고양이'만 AM\_MAMMALIA로 태깅한다. '부엉이카페' 역시 이에 준하여 '부엉이'만 AM\_BIRD로 태깅한다.
      - 고양이 카페 → <고양이: AM\_MAMMALIA> 카페
      - 부엉이 카페 → <부엉이: AM\_BIRD> 카페
- 접사가 결합된 개체명이나 구 구성의 개체명이 사전에 등재되어 있지 않은 경우, 해당 접사나 단어가 결합된 복합어가 사전에 등재되어 있으면 등재된 것에 준하여 태깅한다.
  - 예제
    - 블루베리나무: '블루베리나무'는 사전에 등재되어 있지 않으나, '사과나무', '배나무', '복숭아나무' 등은 등재되어 있으므로 '블루베리나무'도 PT\_TREE로 태깅한다. '블루베리'만 별도로 PT\_FRUIT로 태깅하지 않는다.
    - 피겨스케이팅 선수: '피겨스케이팅 선수'는 사전에 등재되어 있지 않으나, '축구 선수', '야구 선수', '수영 선수' 등은 등재되어 있으므로 '피겨스케이팅 선수'도 CV\_OCCUPATION으로 태깅한다. '피겨스케이팅'만 별도로 CV\_SPORTS로 태깅하지 않는다.
    - 찌개집: 찌개집의 경우 '꽃집'이 사전에 한 단어로 등재되어 있기 때문에 '찌개'만 별도로 내부 태깅을 하지 않는다.

## 라. 태깅 범주

---

### 1. 태깅 범주 결정의 대원칙 - 유일 태깅 원칙

---

- 개체명 중에는 둘 이상의 의미 영역에 동시에 속하여서 둘 이상의 세분류로 태깅이 가능한 것들이 있다. 그러한 경우에 하나의 개체명에 여러 개의 태그를 중복 태깅하지 않고 하나의 태그만을 부여해야 기계적 처리가 명확하고 수월해질 것이다. 따라서 다양한 태그가 부착될 수 있는 개체명 단어가 출현했을 경우에는 태깅 가능한 태그들 중에서 하나의 태그만을 선택하여 할당해야 한다. 이것이 '유일 태깅 원칙'이다.

### 2. 태깅 범주 결정의 원리

---

- 어떠한 개체명이 둘 이상의 세분류로 태깅이 가능하다면, 일차적으로 문맥과 적정 술어를 고려하여 문맥에 맞는 태그를 부착한다. 다만, 문맥으로 판단하는 것이 어려운 경우에는 지침에서 제시하고 있는 우선순위를 적용하여 태깅한다.

#### 2.1. 문맥 우선 적용(원칙)

개체명의 세분류를 판정할 때에 우선적으로 고려해야 하는 것은 문맥이다. 개체명 태깅은 개체명 자체의 세분류를 고려하되, 문장이나 문서 전체에서의 의미를 고려하여 세분류를 최종 판정하고 태깅한다.

- 앞서 언급한 바와 같이 개체명 중에는 문맥에 따라 둘 이상의 의미 영역으로 해석되는 것들이 있다. 예를 들어 식물류, 동물류, 어류 중 일부는 PT 혹은 AM으로 해석되기도 하고 음식 재료로서 CV\_FOOD로 해석되기도 한다. 즉, 문맥에 따라 의미 범주가 달라지는 것이다. 이러한 경우에는 문맥을 우선적으로 적용하여 태깅하되, 문맥으로 판단할 수 없는 경우에만 우선순위를 적용하여 태깅한다.
  - 예제
    - <열빙어: AM\_FISH> 키우기 쉽지 않다.
    - 다음엔 <열빙어: CV\_FOOD> 제철일 때 구워 줘야지.
    - 5월에 심은 <시금치: PT\_GRASS>가 싹이 났다.
    - <시금치: CV\_FOOD> 데쳐서 무쳐 먹으면 얼마나 맛있게요?
  - 둘 이상의 세분류에 속하는 개체명은 해당 세분류의 적정 술어 사용 여부로 판정할 수 있다. 예를 들어, '시금치'가 '심다', '뽑다', '키우다', '자라다', '시들다' 등의 서술어와 함께

사용되었다면, 이때에는 PT\_GRASS로 판정한다. 반면에 '데치다', '볶다', '무치다', '요리하다', '먹다' 등의 서술어와 함께 사용되었다면, 이때에는 CV\_FOOD로 판정한다.

○ 예제

- <부산: LCP\_CITY> <해운대: LCP\_COUNTY> <우동: LCP\_COUNTY>의 한 아파트에서 불이 나 주민 수십 명이 대피하고 1명이 다쳤다.
- <부산: LCP\_CITY> <해운대: LC\_OTHERS> 가 볼 만한 곳 10
- 둘 이상의 범주에 속하는 개체명의 대부분은 문맥과 적정 술어로 세분류를 판정할 수 있다. 예제에서 보인 바와 같이 함께 사용되고 있는 '가 보다' 등의 서술어와 '우동', '아파트', '주민' 등과 같은 명사들을 통해 의미 범주를 판정하고 태깅하면 된다.
- 그러나 서술어 또한 다의성이 있으므로, 적정 술어만으로 세분류를 판정해서는 안 된다. 해당 개체명의 본래의 의미, 문맥, 공기하는 명사 등을 종합적으로 고려하고 서술어의 다양한 의미 중에서 해당 문장에서의 의미가 무엇인지 정확하게 파악하여 신중하게 태깅하도록 한다.

## 2.2. 우선순위 적용(예외)

- 문맥을 통해 해당 개체명의 의미 범주를 판별할 수 없는 경우에 한하여, 아래에 제시한 우선순위를 적용하여 태깅한다.

예제	기능 태그 1	기능 태그 2
미니미니, 뽀뽀이 등	PS_NAME	PS_CHARACTER/PS_PET
청와대, 국립중앙박물관 등	OG 계열	AF_BUILDING
미아사거리 등	AF_ROAD	LC_OTHERS
한국, 프랑스 등	LCP_COUNTRY	OGG_POLITICS
고양시, 성남시 등	LCP_CITY	OGG_POLITICS
혜화, 시부야 등(동일한 역 이름)	LCP_COUNTY	LC_OTHERS
시베리아 등(동일한 역 이름)	LCG_CONTINENT	LC_OTHERS
제주도 등	LCP_PROVINCE	LCG_ISLAND
닭, 오리, 고등어 등	AM 계열	CV_FOOD
바이러스 등	AM_OTHERS	TMM_DISEASE
시금치, 대추, 옥수수 등	PT 계열	CV_FOOD
크롬, 칼륨, 티타늄 등	MT_METAL	MT_ELEMENT
산소, 질소, 규소 등	MT_ELEMENT	MT_CHEMICAL

- 위의 예제들은 태그1과 태그2로 모두 태깅할 수 있는 것들이다. 문맥을 통해 해당 개체명의 의미 범주를 판별할 수 없다면, 굵게 표시된 가능 태그1로 우선 적용하여 태깅하면 된다. 여기에서는 대표적인 사례만 제시한 것이며, 자세한 설명과 예문은 '바. 세부분류 개체명 정의'의 각 항목에서 확인할 수 있다.

## 마. 특수 항목 태깅

---

### 1. 괄호 안의 원어 정보

---

- 원어 정보가 개체명 사이에 괄호로 일부만 제시된 경우는 괄호를 제외하고 앞뒤를 합쳐야만 개체명이 되기 때문에 전체를 묶어 괄호까지 포함하여 태깅한다.
  - 예제
    - 퓨(Pew)공익재단의 → <퓨(Pew)공익재단: OGG\_OTHERS>의
    - 테링(鐵嶺)시 → <테링(鐵嶺)시: LCP\_CITY>
- '이코노미스트지(誌)'와 '윤덕여호(號)'처럼 원어 정보가 개체명 끝부분에 괄호로 일부만 나왔을 때는 괄호 정보를 제외해도 하나의 개체명으로 인식이 가능하기 때문에 괄호정보를 태그하지 않는다.
  - 예제
    - 이코노미스트지(誌)는 → <이코노미스트지: OGG\_MEDIA>(誌)는
    - 윤덕여호(號) → <윤덕여호: OGG\_SPORTS>(號)
- '한국문화예술위원회(이하 문예위)', 미국 드라마(미드)'와 같이 괄호 정보가 해당 개체명의 줄임말일 경우, 해당 텍스트에서 괄호 안의 정보가 단독으로 나와 개체명으로 인식해야 하기 때문에 생략하거나 통합하지 않고 각각 태깅한다.
  - 예제
    - 한국문화예술위원회(이하 문예위) → <한국문화예술위원회: OGG\_ART>(이하 <문예위: OGG\_ART>)
    - 미국 드라마(미드) → <미국 드라마: CV\_ART>(<미드: CV\_ART>)
- 주의
  - 개체명과 함께 괄호 정보가 개체명 원어 정보 전체를 나타낼 경우에는 괄호 정보도 따로 태깅한다.
  - 후진타오(胡錦濤) → <후진타오: PS\_NAME>(<胡錦濤: PS\_NAME>)
  - 수월관음도(水月觀音圖) → <수월관음도: AFA\_ART\_CRAFT>(<水月觀音圖: AFA\_ART\_CRAFT>)

### 2. 괄호 안의 부가 정보

---

- 괄호 안의 부가 정보가 개체명일 경우에는 괄호 안의 정보도 태깅한다.
  - 예제
  - 이번 불화(세로 123.5cm×가로 60.5cm) → 이번 <불화: CV\_ART>(<세로 123.5cm×가로 60.5cm: QT\_OTHERS>)

- '이름(국가, 순위, 나이)'로 나온 경우 괄호 안의 정보를 각각 '국가: LCP\_COUNTRY, 순위: QT\_ORDER, 나이: QT\_AGE'로 태깅한다.
  - 예제
    - 샤라포바(러시아, 2위, 28세) → <샤라포바: PS\_NAME>(<러시아: LCP\_COUNTRY>, <2위: QT\_ORDER>, <28세: QT\_AGE>)
- 주의
  - 단, 괄호 안의 부가 정보가 개체명 태깅 범위에 속하지 않는 경우에는 괄호 정보 앞의 개체명만 태깅한다.
  - 가운뎃점이나 다른 기호가 개체명 안에 포함되어 있는 경우 함께 태깅한다.
    - 5·8시 → <5·8시: TI\_HOUR> 취득·등록세 → <취득·등록세: CV\_TAX>
  - 괄호 정보 속 기호 'x'도 개체명 내부에 있는 경우 함께 태깅한다.

### 3. 원어절 사이에 기호가 삽입되어 있는 경우

- '이천 이~ 십년'과 같이 중간에 '~' 등의 기호가 삽입되어 있는 경우에는 기호를 포함하여 태깅한다.
- '~' 등의 기호는 형태 분석 시에 담화표지로 인식된다. 따라서 포함하여 태깅하되, 의미가 없는 것으로 판단한다.
  - 예제
    - 이천 이~ 십년 오월과 칠월 → <이천 이~ 십년: DT\_YEAR> <오월: DT\_MONTH>과 <칠월: DT\_MONTH>
    - 바-나-나 → <바-나-나: PT\_FRUIT>

### 4. 날짜/시간/수량 N + 보조사/접사/단위성 의존명사

- 날짜, 수량, 시간 표현이 보조사, 접사, 단위성 의존명사 등과 함께 쓰인 경우 통합하여 함께 태깅한다. 따라서 '-만, -쯤, -씩, -간, -경'이나 '전, 후, 뒤, 초' 등도 포함하여 태깅한다.
  - 예제
  - DT
    - 100년간: DT\_DURATION
    - 9월쯤: DT\_MONTH
    - 2002년경: DT\_YEAR
    - 100년 전: DT\_OTHERS
    - 20년 후: DT\_OTHERS
    - 90년대 초: DT\_YEAR



- TI
  - 6시쯤: TI\_HOUR
  - 3시간 후: TI\_OTHERS
  - 오전 10시 전까지: TI\_OTHERS
  - 7시쯤까지: TI\_OTHERS
- QT
  - 16살까지: QT\_AGE
  - 14세 이상부터: QT\_AGE
  - 10여 명: QT\_MAN\_COUNT
- 주의
  - 날짜/수량/시간 N + 보통명사: 정도, 가량, 이상, 이하, 이내 등과 같은 경우도 포함하여 태깅한다.
    - 0.5초 정도: TI\_SECOND
    - 일주일 가량: DT\_DURATION
    - 17세 이상: QT\_AGE
    - 4개월령 이상: QT\_AGE
    - 약 320km 이하: QT\_LENGTH
    - 한두 가지 정도: QT\_COUNT
  - '3일 연속' 등에서의 '연속'은 태깅하지 않는다.

## 5. 영어나 한자 등의 외국어

- 문장 안에서 원어로만 쓰인 '인명, 지명, 기관명, 상품명, 작품명'은 태깅하되, 문장 혹은 단락 전체가 외국어로 쓰인 경우에는 일괄적으로 태깅하지 않는다. 외국어의 경우 고유명만 태깅하고 보통명사는 태깅하지 않는다.
  - 예제
    - (인명) Post Malone: PS\_NAME
    - (지명) PARIS: LCP\_CAPITALCITY
    - (기관명) Starbucks: OGG\_FOOD
    - (상품명) iPhone11pro: AFW\_OTHER\_PRODUCTS
    - (작품명) The Devil Wears Prada: AFA\_VIDEO
- 괄호 안에 원어 정보로 추가되어 있는 경우, 별도로 태깅한다.
  - 예제
    - 본초강목(本草綱目) → <본초강목: AFA\_DOCUMENT>(<本草綱目: AFA\_DOCUMENT>)
    - 뉴욕시(New York City) → <뉴욕시: LCP\_CITY>(<New York City: LCP\_CITY>)

## 6. 줄임말

- 단어의 일부분을 줄이거나 여러 단어를 한 단어로 줄여 만든 줄임말의 경우 원래의 단어에 해당하는 개체명으로 태깅한다.
  - 예제
    - 포메 → <포메: AM\_MAMMALIA> (포메라니안)
    - 라흐 → <라흐: PS\_NAME> (라흐마니노프)
    - 카톡 → <카톡: TMI\_SERVICE> (카카오톡)
    - 스벅 바닐라라떼가 제일 맛있어. → <스벅: OGG\_FOOD> (스타벅스)
    - 우행시 보러 가자. → <우행시: AFA\_VIDEO> (우리들의 행복한 시간)

## 7. 개체명의 일부로 해당 개체를 가리키는 경우

- 문맥을 고려했을 때 특정 개체명의 일부로 해당 개체를 가리킨다고 판단되는 경우, 복원된 형태에 준하여 문맥의 의미에 해당하는 세분류로 태깅한다.
  - 예제
    - '수원 삼성' (중략) 수원이 3대 0으로 울산을 이겼다.
      - '수원'은 바로 앞의 '수원 삼성'을 지칭하며, '이기다'와 함께 사용된 것을 고려하여 'LCP\_CITY'가 아니라 'OGG\_SPORTS'로 태깅한다.
    - <부산국제영화제: EV\_FESTIVAL> 개막. '엑소 수호' 배우 김준면으로 <부산: EV\_FESTIVAL> 레드카펫 밟는다.
      - 이 문장에서 '부산'은 도시를 지칭하는 것이 아니라 바로 앞에 나온 '부산국제영화제'를 지칭하는 것이므로 'EV\_FESTIVAL'로 태깅한다.
    - 내가 상당히 어려워하는 작곡가 <모차르트: PS\_NAME> <피아노 소나타: AFA\_MUSIC>. 정말 <모차르트: AFA\_MUSIC> 너무 어렵다.
      - 두 번째 문장에서 '모차르트'는 작곡가 '모차르트'가 아니라 바로 앞의 문장에 있는 '모차르트의 피아노 소나타'를 지칭하고 있기 때문에 여기에서 '모차르트'는 'AFA\_MUSIC'으로 태깅한다.
  - 상위어, 혹은 보통명사로 포괄적으로 지칭하는 경우에는 태깅하지 않는다.
    - <보듬 H형 하네스: AFW\_OTHER\_PRODUCTS>를 구입하기로 했다. H형 하네스를 사면서 리드줄도 함께 구입했다.
      - 'H형 하네스'는 보통명사로 포괄적으로 지칭하는 것이므로 태깅하지 않는다.
    - <서울대학교: OGG\_EDUCATION>가 다음 학기 등록금을 동결하기로 했다. 대학교 측은 코로나로 인한 경제 위축과 비대면 수업의 연장 가능성으로 이 같은 결정을 한 것으로 알려졌다.

- '대학교'는 바로 앞 문장의 '서울대학교'를 지칭하나, 상위어를 통한 대명사적 쓰임이므로 태깅하지 않는다.
- 즉, 개체명의 일부로 해당 개체를 가리키는 경우에는 복원된 형태에 준하여 문맥의 의미에 해당하는 세분류로 태깅하나, 상위어 혹은 보통명사인 경우에는 태깅하지 않는다.

## 8. 비유적 표현

---

- 비유적 표현이라 할지라도 비유인지 아닌지를 고려하지 않고 해당 개체명 분류에 알맞게 태깅한다. (원래의 의미에 따라)
  - 예제
    - 민주당의 입장에서는 '않던 이'가 빠진 격이다. → 민주당의 입장에서는 '않던 <이: AM\_PART>'가 빠진 격이다.
    - 그는 결혼과 취직이라는 '두 마리 토끼'를 동시에 잡았다. → 그는 결혼과 취직이라는 '<두 마리: QT\_COUNT> <토끼: AM\_MAMMALIA>'를 동시에 잡았다.

## 9. 개체명과 조사/어미가 하나의 음절로 실현된 경우

---

- 개체명과 어미 혹은 개체명과 조사가 하나의 음절로 실현된 경우, 어형 수정 후 형태소가 제거된 개체명만을 태깅한다.
  - 예제
    - 삼십프로인가? → <삼십프로: QT\_PERCENTAGE>인가?
    - 러시아인데요. → <러시아: LCP\_COUNTRY>인데요.
    - 세 겔 → <세 개: QT\_COUNT>를

## 10. 오자나 오타

---

- 원시어절에서 명백한 오타(전사상 오류, 발음 오류 등)로 보이는 경우에도 개체명 분류에 속하면 원어절 그대로 해당 개체명으로 태깅한다.
- 원문을 수정해서 태깅하지 않는다. (후처리로도 원문을 수정하지 않음.)
  - 예제
    - 오스트릴라에 → <오스트릴라: LCP\_COUNTRY>에
    - 사라코지 대통령이 → <사라코지: PS\_NAME> <대통령: CV\_POSITION>이

## 바. 세부분류 개체명 정의

---

### 1. PERSON

---

- PERSON은 "인명 및 인물의 별칭"을 나타낸다.
- 정확히 사람을 나타내는 개체는 아닐지라도, 사람과 유사한 지위를 가진다고 판단되는 캐릭터명, 반려동물 이름 등의 "유사 인물 명칭" 또한 PERSON의 범주 안에 포함한다.

#### 1.1. PS\_NAME

##### 1.1.1. 정의 및 설명

- 인명 및 인물의 별칭
  - 인명의 하위 유형
  - 실존 인물의 이름
  - 비실존 인물의 이름
  - 별칭

##### 1.1.2. 분석 대상

- 실존 인물의 이름
  - 역사 인물의 이름: 흥선대원군, 혜원 신윤복, 현종, 조조, 유비, 손권, 프랭클린 루스벨트, 콘스탄티누스, 루이 14세, W.가이거, S.A.키르케고르, F.M.도스토예프스키, 마르쿠스 아우렐리우스
  - 가수, 배우, 작가 등의 활동명: 아이유, 싸이, 공유, 하정우, 김수로, 이상, 기안84
  - 가수 그룹의 이름: 방탄소년단, 소녀시대, 블랙핑크, 여자친구
  - 배우의 이름: 원빈, 한가인, 장동건, 짐 캐리
  - 감독의 이름: 봉준호, 박찬욱, 스티븐 스필버그
  - 그 밖에 일반인, 유명인 등 모든 인물의 이름
- 비실존 인물(신화 및 종교 인물)의 이름
  - 신화 등장인물의 이름: 단군, 웅녀, 전륜성왕, 신농, 아누비스, 오시리스, 제우스, 아킬레스, 티에스테스, 아가멤논, 엔키두
  - 종교 인물의 이름: 부처, 예수, 유다, 보현보살

- 별칭
  - 별명: 척척박사, 돌부처, 떠벌이
  - 태명: 대박이, 튼튼이, 건강이
  - 필명: 김수현, 이말년, 기안84, 조지 오웰, 조지 엘리엇, 메리 웨스트매컷
  - 세레명: 요한, 가브리엘, 클라라, 빅토르
  - 인터넷 사용자명 및 닉네임: 엠브로, 썬양, 로아맘, 페이커, 엘란, 용개, 맑고흐림
  - 인터넷포털 및 SNS 계정명(ID): @callmegray, @SAMSUNGKOREA, @jtbcdrama, sojujsh, mcelan, sujin0905

### 1.1.3. 예문

- 실존 인물의 이름
  - 또한 <쿤: PS\_NAME>은 <S.A.키르케고르: PS\_NAME>나 <M.하이데거: PS\_NAME>의 실존론적(實存論的) 사고방식에 따라 미적 기분을 다루었다.
  - 한편 <키르케고르: PS\_NAME>는 데모니세를 죄악감에 빠진 인간이 악에서 헤어날 수 없는 채 선에 대하여 품는 불안감이라고 설명하였다.
  - 1902년 독일 유학에서 돌아온, <H.스테펜스: PS\_NAME>가 독일 낭만주의 정신을 고취하게 되자 <A.월렌슈레거: PS\_NAME>를 선두로 낭만주의 운동이 일어나, <N.그룬트비: PS\_NAME> · <안데르센: PS\_NAME> · <키르케고르: PS\_NAME> 등이 이어받았다.
  - 1942년 영국에서 출생한 물리학자 <스티븐 호킹: PS\_NAME> 박사는 2018년 3월, 영원히 우리의 곁을 떠나고 말았다.
  - <레드벨벳: PS\_NAME>이 새 앨범으로 돌아왔다.
  - <세시봉: PS\_NAME>의 <윤형주: PS\_NAME> 장로님이 함께한 행복축제, 정말로 너무 좋았다.
  - '살인의 추억', '괴물' 등을 연출한 한국의 거장 <봉준호: PS\_NAME> 감독은 '기생충'을 통해 한국 영화로는 최초로 칸 영화제 황금종려상을 수상해 냈다.
- 비실존 인물(신화 및 종교 인물)의 이름
  - <엔키두: PS\_NAME>는 <길가메시: PS\_NAME>에게 안긴 채 죽었다.
  - <제우스: PS\_NAME>는 그리스 신화의 최고신으로 "신들과 인간들의 아버지"로 불린다.
  - 겸허와 관용, 그리고 무욕. 우리는 <부처: PS\_NAME>의 이 가르침을 잊지 말아야 할 것이다.
- 별칭
  - 시청자들에게 큰 사랑을 받았던 <대박이: PS\_NAME>가 벌써 7살이 됐다고 한다.
  - '동물농장', '1984' 등으로 유명한 <조지 오웰: PS\_NAME>은 사실 그의 진짜 이름이 아니다.
  - 해당 포스팅의 저작권은 <엘란: PS\_NAME>에게 있으니 무단 불펌을 금지합니다.

- 당첨자는 <@웃그제: PS\_NAME> 이웃님이시네요.

#### 1.1.4. 주의 사항

- (받침이 있는 이름+접미사 '-이') 받침이 있는 형태로 끝나는 이름 뒤에 어조를 고르는 접미사 '-이'가 붙어 쓰인 경우 이를 포함한 전체를 PS\_NAME으로 태깅한다.
  - 영숙이는 집으로 돌아갔다. → <영숙이: PS\_NAME>는 집으로 돌아갔다.
  - 나이를 먹어도 영진이는 여전히 철이 없다. → 나이를 먹어도 <영진이: PS\_NAME>는 여전히 철이 없다.
- (활동명) 유명인이 활동 당시 공식적으로 사용한 이름은 PS\_NAME으로 태깅한다.
  - 가수 양파가 엄청난 가창력을 뽐내고 있다. → 가수 <양파: PS\_NAME>가 엄청난 가창력을 뽐내고 있다.
  - 조세호는 양배추라는 활동명을 버리고 본명 조세호로 돌아왔다. → <조세호: PS\_NAME>는 <양배추: PS\_NAME>라는 활동명을 버리고 본명 <조세호: PS\_NAME>로 돌아왔다.
- (별명, 별칭, 애칭) 별명이나 별칭 등 비유적인 표현이 특정한 한 개인을 지칭하는 것이 명확하다면 PS\_NAME으로 태깅한다.
  - 베이브류스 류현진 완벽 투구 → <베이브류스: PS\_NAME> <류현진: PS\_NAME> 완벽 투구
  - 내 친구 철수는 행동이 너무 느려 나무늘보로 통한다. → 내 친구 <철수: PS\_NAME>는 행동이 너무 느려 나무늘보<PS\_NAME>로 통한다.
  - 로이맘 님이 아기 용품을 나눔해 주셨어요. → <로이맘: PS\_NAME> 님이 아기 용품을 나눔해 주셨어요.
  - 두 개의 심장 박지성은 결코 지치지 않았다. → 두 개의 심장 <박지성: PS\_NAME>은 결코 지치지 않았다.
  - '두 개의 심장'이라는 별명은 '박지성'이라는 한 개인만을 지칭하는 것이 아니라, '열심히 뛰는 운동선수'를 모두 가리키는 포괄적인 의미를 가지고 있으므로 PS\_NAME으로 태깅하지 않는다.
- (성+씨) 특정한 개인을 지시하는 것이 명확한 경우에는 '성'만 PS\_NAME으로 태깅한다. 특정인이 아닌 가문을 의미할 때는 태깅하지 않는다.
  - 김 씨, 이것 좀 봐. → <김: PS\_NAME> 씨, 이것 좀 봐.
  - 민 씨는 조선 시대 명문가 중 하나이다. → 민 씨는 조선 시대 명문가 중 하나이다.
    - 이 때의 '민'은 태깅하지 않는다.
- (호+이름) '호+이름' 구성의 경우 결합된 표현이 한 대상을 지칭하므로 전체를 PS\_NAME으로 태깅한다. 또한 문서 내에서 호가 단독으로 출현한 경우에도 이를 PS\_NAME으로 태깅한다. 단, 묘호와 이름 사이에 직위나 직책명이 삽입되어 두 개체명 사이의 명확한 변별이 이루어진 경우에 한정하여 호, 직위명, 본명을 각각 태깅하도록 한다.

- <퇴계 이황: PS\_NAME>은 조선 중기의 학자이자 문신이다.
- <추사: PS\_NAME>의 글씨를 보고 있노라면 형용할 수 없는 경외감이 싹트곤 한다.
- <정조: PS\_NAME> <대왕: CV\_POSITION> <이산: PS\_NAME>은 조선의 제22대 왕이다.
- (커플명) 커플 이름처럼 두 사람을 하나로 묶어 지칭하는 표현들 중, 개별적인 형태로 복원이 가능한 경우에는 그 각각을 PS\_NAME으로 태깅한다.
  - 드라마 태양의 후예에서 송중기와 송혜교 커플을 두고 나온 신조어 <송: PS\_NAME> <송: PS\_NAME> 커플
- (인명+지위, 지위+인명, 인명+호칭) 다음과 같은 명사구는 지위, 인명, 호칭 따위를 각각 따로 태깅한다.
  - 문재인 대통령 → <문재인: PS\_NAME> <대통령: CV\_POSITIOIN>
  - 이재용 부회장 → <이재용: PS\_NAME> <부회장: CV\_POSITIOIN>
  - 국대비 왕씨 → <국대비: CV\_POSITIOIN> <왕: PS\_NAME> 씨
  - 폐비 윤씨 → <폐비: CV\_POSITIOIN> <윤: PS\_NAME> 씨
  - 김 박사 → <김: PS\_NAME> <박사: CV\_POSITIOIN>
- 단, 상기 명사구가 하나의 인명으로 굳어져 사용되는 경우에는 전체를 PS\_NAME으로 태깅한다.
  - 신바람 이박사 → 신바람 <이박사: PS\_NAME>
  - 가수 김씨는 2013년 이혼했다. → 가수 <김씨: PS\_NAME>는 2013년 이혼했다.
- (작곡자명) '작곡자명+연주곡명'과 같은 명사구의 경우 작곡자는 PS\_NAME으로, 연주곡명은 AFA\_MUSIC으로 태깅한다.
  - 다음으로 쇼팽 Op10-12를 들려 드리겠습니다. → 다음으로 <쇼팽: PS\_NAME> <Op10-12: AFA\_MUSIC>를 들려 드리겠습니다.
- (인명 중첩) 동일 음절 중복으로 두 인명이 이어진 경우, 전체가 특정 인물을 지칭하므로 전체를 하나로 묶어 PS\_NAME으로 태깅한다.
  - 이명박근혜 → <이명박근혜: PS\_NAME>
- 구문 맥락상 세부 항목 구별이 명확하지 않은 경우에는 PS\_NAME을 우선적으로 태깅한다.
  - 내 사랑 <미니미니: PS\_NAME>
  - 우리 <뽀뽀이: PS\_NAME>, 엄마가 미안해.
  - 나에게 큰 희망과 용기를 준 <김지영: PS\_NAME>. 고마워.

## 1.2. PS\_CHARACTER

- PS\_CHARACTER는 주로 소설이나 드라마, 만화영화, 게임 등 가상의 세계에 등장하는 인위적으로 가공된 가상 인물을 뜻한다.
- 독특한 인물이나 동물의 모습을 디자인에 도입한 캐릭터 상품과 이것이 one source multi-use 과정(하나의 소스로 여러 종류의 상품 유형으로 발전시키는 과정)을 통해 다른

장르에 파생된 경우 역시 PS\_CHARACTER에 포함된다.

- 동물, 식물, 사물 등이 의인화 과정을 통해 사람에 준하는 개체로 판단되는 경우에도 PS\_CHARACTER로 태깅한다.

### 1.2.1. 정의 및 설명

- 가상 캐릭터 이름
- 작품 속 등장인물의 이름

### 1.2.2. 분석 대상

- 가상 캐릭터 이름
  - 만화영화 등장인물의 이름: 뽀로로, 루피, 짱구, 심슨, 심바, 푸, 미키마우스, 라푼젤, 우디, 버즈, 배트맨, 슈퍼맨, 아이언맨, 원더우먼, 캡틴 마블, 블랙 팬서
  - 게임 캐릭터의 이름: 마리오, 소닉, 팩맨, 류, 우디, 제라툴, 가렌, 아이번, 하이머딩거
  - 상품 캐릭터의 이름: 펑수, 뿌까, 어피치, 프로도, 무지, 라이언, 브라운, 샬리
- 작품 등장인물의 이름
  - 소설 등장인물의 이름: 김지영, 장발장, 해리 포터, 오베, 쇼타
  - 드라마 등장인물의 이름: 김삼순, 유시진, 박태수, 홍라온, 존 스노우
  - 영화 등장인물의 이름: 인디애나 존스, 제임스 본드, 피터 파커, 클라크 켄트

### 1.2.3. 예문

- 가상 캐릭터 이름
  - 이상하게도 요 근래 우리나라는 펑귄 캐릭터가 인기가 많다. <뽀로로: PS\_CHARACTER>로 시작해서 <펑수: PS\_CHARACTER>까지 이어지는 펑귄 대박 행진을 보라.
  - 라인프렌즈를 대표하는 캐릭터 <브라운: PS\_CHARACTER>은 일본에서 특히 인기가 많다.
  - <소닉: PS\_CHARACTER>은 고슴도치를 모티브로 만들어졌다.
- 작품 등장인물의 이름
  - <피터 파커: PS\_CHARACTER>가 바로 <스파이더맨: PS\_CHARACTER>이다.
  - <헐크: PS\_CHARACTER>는 평범한 인간의 힘을 한참이나 초월한 존재이지.
  - 영화 '레미제라블'에서 <장발장: PS\_CHARACTER>은 빵 한 조각을 훔쳐 5년의 징역을 살았다.
  - <존 스노우: PS\_CHARACTER>를 연기한 사람의 이름이 뭐였지?
  - <제임스 본드: PS\_CHARACTER>란 이름은 007시리즈의 상징과도 같다.
  - 허영만의 만화 타짜는 <김곤: PS\_CHARACTER>이라는 인물을 통해 도박과 그에 얽힌 인간들의 군상을 적나라하게 드러낸다.



- 알라딘에서 <자스민: PS\_CHARACTER> 공주를 연기한 배우 나옴이 스콧이 내달 방한한다.
- 디즈니의 1993년작 알라딘이 실사화를 결정했다. 퍼런 피부와 유쾌한 표정이 트레이드마크인 <알라딘: PS\_CHARACTER>을 과연 누가, 어떻게 소화할 수 있을지 기대된다.

#### 1.2.4. 주의 사항

- (작품 속 실존 인물) 소설이나 영화, 드라마 등 작품에 등장하는 실존 인물은, 구문 맥락상 '작품 속의 인물'을 지칭하는 것이 명확하다면 PS\_CHARACTER로 태깅한다.
  - 영화 '잡스'에서 <스티브 잡스: PS\_CHARACTER> 역을 맡은 애쉬튼 커처가 팬들을 향해 손을 흔들고 있다.
  - 영화는 숨을 거두기 직전의 <마리 퀴리: PS\_CHARACTER>의 모습을 보이며 시작한다.
  - 배우 최수종은 "20년 전 출현했던 드라마 한 편에서의 강렬했던 이미지 때문인지 지금까지도 <왕간: PS\_CHARACTER> 아저씨라는 호칭을 자주 듣는다"며 웃어 보였다.
- (캐릭터의 별칭) PS\_CHARACTER 개체를 캐릭터명이 아닌 또 다른 별칭이나 별명으로 부르는 경우, 이 역시 PS\_CHARACTER로 태깅한다.
  - 아니 왜 집어치우시죠? 우리집에도 <곰빠까: PS\_CHARACTER> 걸어 주세요.
  - <가봉아: PS\_CHARACTER>가 가렌 놀리는 말인건 아는데 가렌이랑 무슨 단어가 합쳐진 건가요?
- (마스코트) TV프로그램, 기업, 스포츠팀 등에서 공인한 공식적인 마스코트는 PS\_CHARACTER로 태깅한다.
  - <블레오: PS\_CHARACTER>는 삼성 라이온즈의 마스코트이다.
  - 추억의 대전엑스포 마스코트 <꿈돌이: PS\_CHARACTER>, 예능에서 만나요.
  - 텔레비 오사카 '와후소우혼케'의 마스코트 강아지 <마메스케: PS\_CHARACTER>는 기수가 있다는 점에서 특별하다.
- 구문 맥락상 지칭 대상이 특정한 개인 가상 인물이 확실한 경우에만 PS\_CHARACTER로 태깅한다.
  - 김지영 보러 영화관에 갔는데, 김지영이 한 말이 너무나도 와 닿더라. → <김지영: AFA\_VIDEO> 보러 영화관에 갔는데, <김지영: PS\_CHARACTER>이 한 말이 너무나도 와 닿더라.
  - 해리포터 시리즈는 전세계 5억만 부 이상 판매된 세기의 베스트셀러이다. 소설은 해리포터가 마법학교에 입학하며 겪게 되는 모험과 환상을 그리고 있다. → <해리포터: AFA\_DOCUMENT> 시리즈는 전세계 5억만 부 이상 판매된 세기의 베스트셀러이다. 소설은 <해리포터: PS\_CHARACTER>가 마법학교에 입학하며 겪게 되는 모험과 환상을 그리고 있다.

### 1.3. PS\_PET

- PS\_PET의 태깅 대상인 '반려동물의 이름'을 일반적인 동물명인 ANIMAL과는 다른 개체로 판단하는 기준은 다음과 같다.
  - 첫째, '반려동물'은 동물이 사람의 장난감 같은 존재가 아니라 사람과 더불어 살아가는 반려자라는 인식이 더해진 표현이다.
  - 둘째, '반려동물'로서 인간과 더불어 생활하는 동물은 일반적인 가축이나 야생동물과는 달리 인식된다.
  - 셋째, 결국 사람이 반려동물에 이름을 붙여 부르는 행위는 반려동물을 사람과 동등한 위치의 '친구'나 '가족'으로 인식한다는 것과 같다.
- 위와 같은 근거에 따라, 사람이 반려동물에게 특정한 이름을 붙여 부르는 경우에도, 그 동물을 사람에 준하는 개체로 인정했다고 보고 PS\_PET으로 태깅한다.

#### 1.3.1. 정의 및 설명

- 반려동물의 이름

#### 1.3.2. 분석 대상

- 반려동물 이름: 뽀빠, 쿠키, 바둑이, 까망이, 나비, 까미, 루피, 초코, 탄이

#### 1.3.3. 예문

- 반려동물 이름
  - 우리 <쿠키: PS\_PET>는 낯선 사람을 봐도 꼬리부터 흔들고 본다.
  - <까미: PS\_PET>는 산책을 좋아하는데 <레미: PS\_PET>는 죽기보다 싫어한다.
  - 야생 본능을 잊고 '개냥이'처럼 사람을 따르는 삶 <나비: PS\_PET>의 이야기가 전해졌다.
  - 저 녀석이 우리 사파리의 서열 1위인 <심바: PS\_PET>입니다.

#### 1.3.4. 주의 사항

- (유아어) 주로 어린아이들이 동물을 지칭할 때 사용하는 유아어 '야옹이', '멍멍이', '꿀꿀이' 등은 실제 애완동물의 이름이 아닌 이상 PS\_PET이 아닌 AM으로 태깅한다.
  - <야옹이: AM\_MAMMALIA>는 야옹야옹하고 울어요.
- (동물의 별칭) 온라인상에서 주로 사용되는 개, 고양이 등의 별칭인 '댕댕이', '냥이' 등도 실제 애완동물의 이름이 아닌 이상 PS\_PET이 아닌 AM으로 태깅한다.
  - 수진이네 <댕댕이: AM\_MAMMALIA>가 무슨 종이였더라? 리트리버였나?
  - 박신혜, 반려묘와 행복한 일상. <냥이: AM\_MAMMALIA>는 집사가 박신혜라 좋겠네.

## 2. STUDY\_FIELD

STUDY\_FIELD는 학문 분야를 나타낸다. 세부 범주로는 FD\_SCIENCE, FD\_SOCIAL\_SCIENCE, FD\_MEDICINE, FD\_ART, FD\_HUMANITIES 다섯 가지로 구성되어 있다. 이외의 학문 분야는 FD\_OTHERS에 해당된다. 더불어 STUDY\_FIELD는 학파를 포함한다. 과학, 사회과학, 의학, 예술, 인문학 관련 학파 및 유파들은 STUDY\_FIELD의 세부 범주에 해당된다.

- 주의
  - 특정 기관의 세부 조직으로서의 '-학과'는 태깅 대상이 아니다. 학문 분야로서의 '-학'만 STUDY\_FIELD로 태깅한다.
  - 학파는 STUDY\_FIELD로 태깅하며, 이론, 사조, ~주의는 THEORY로 태깅한다. 그러나 '학파+이론'과 같이 나올 경우, 최장 단위 태깅 원칙에 따라 '학파+이론' 전체를 THEORY로 태깅한다.
    - <신고전학파 성장론: TR\_SOCIAL\_SCIENCE>의 경우에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 <신고전파종합(neoclassical synthesis): TR\_SOCIAL\_SCIENCE>의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.
    - <신고전학파: FD\_SOCIAL\_SCIENCE> <성장론: TR\_SOCIAL\_SCIENCE>으로 태깅하지 않는다.
  - STUDY\_FIELD의 범위는 한국연구재단(NRF)의 학술연구분야분류에서 대분류, 중분류, 소분류에 해당되는 것이다. 세분류에 해당하는 것은 STUDY\_FIELD로 태깅하지 않는다.
    - <국어국문학: FD\_HUMANITIES>을 공부하면서 접한 고전소설이 잊히지 않는다.
    - '국어국문학'은 한국연구재단의 학술연구분야분류에서 중분류에 해당하므로 STUDY\_FIELD의 하나로 태깅하지만 '고전소설'은 세분류에 해당하므로 STUDY\_FIELD로 태깅하지 않는다. '고전소설'은 CV\_ART로 태깅한다.
    - '국어국문학'은 한국연구재단의 분류에서 '인문학'에 속하는 것으로 되어 있으므로 FD\_HUMANITIES로 태깅한다.

### 2.1. FD\_SCIENCE

#### 2.1.1. 정의 및 설명

- 과학 관련 학문 분야 및 과학 학파
- FD\_SCIENCE의 하위 유형
  - 순수과학(자연과학): 물리학, 생명과학, 수학, 지구과학, 화학 등
  - 기술과학(응용과학): 건축공학, 금속재료공학, 기계공학, 무기 및 병기공학, 산업공학, 수

의학, 원자력공학, 자동차공학, 전기공학, 조선해양공학, 토목건설공학, 항공우주공학, 화학공학, 환경공학 등

### 2.1.2. 분석 대상

- 순수과학
  - 물리학, 생명과학, 수학, 지구과학, 화학, 생물물리화학, 소립자물리학, 뉴턴역학, 양자역학, 생물학, 유전학, 응용지질학, 천문학 등
- 기술과학
  - 건축공학, 금속재료공학, 기계공학, 뇌공학, 무기 및 병기공학, 산업공학, 수의학, 원자력공학, 자동차공학, 전기공학, 조선해양공학, 토목건설공학, 항공우주공학, 화학공학, 환경공학 등

### 2.1.3. 예문

- 순수과학
  - <생물학: FD\_SCIENCE>은 생물의 종류·구조·기능, 그 밖의 연구 대상에 따라 여러 가지 분야로 나눌 수 있지만,
  - <유전학: FD\_SCIENCE>은 생물의 형질 발현에 대한 원인을 구명하는 과학으로
  - <응용지질학: FD\_SCIENCE>은 옛날에는 주로 광물자원의 생성과정과 성인(成因) 및 경제성과 분포에 관해 연구하는 광상학(鑛床學: economic geology)으로 대표되었다.
  - 고등학교 교육 과정의 <지구과학: FD\_SCIENCE>은 이와 같은 넓은 뜻의 지구과학을 가리킨다.
- 기술과학
  - 그래서 <뇌과학: FD\_SCIENCE>이 우리가 궁금하고 또 흥미가 있고 앞으로 우리가 연구해야 될 것이다, 이런 생각을 하는데 아까 잠깐 제가 말씀드렸지만 <뇌과학: FD\_SCIENCE>은 그렇게 한 사람이, 한 분야가 할 수가 있는 게 아니죠.
  - DNA와 혈청, 혈장 등 각종 질병과 백신 개발 등 우리나라 <생명공학: FD\_SCIENCE> 연구의 핵심 생체자원을 실은 바이오뱅크입니다.

### 2.1.4. 주의 사항

- '과학, 물리'와 같은 넓은 범위부터 '뉴턴역학'과 같은 세부 학문까지 태깅한다. '상대성 원리', '케플러 법칙'과 같은 과학 이론은 TR\_SCIENCE로 태깅하지만 '뉴턴역학'과 같은 이론들이 정리된 하나의 학문은 FD\_SCIENCE로 태깅한다.
  - 그 다음에 내가 물리를 했어요. → 그 다음에 내가 <물리: FD\_SCIENCE>를 했어요.
  - 케플러 법칙과 뉴턴의 법칙 등에 대해 다루는 뉴턴역학은 자연과학, 공학 등 다양한 학문에 영향을 주었다. → 케플러 법칙과 뉴턴의 법칙 등에 대해 다루는 <뉴턴역학: FD\_SCIENCE>은 <자연과학: FD\_SCIENCE>, <공학: FD\_SCIENCE> 등 다양한 학문에 영향

을 주었다.

- '케플러 법칙'과 '뉴턴의 법칙'은 각각 <TR\_SCIENCE>로 태깅한다.
- '수의학'은 FD\_MEDICINE이 아니라 FD\_SCIENCE로 태깅한다. FD\_MEDICINE은 인간을 질병으로부터 구하고 건강법을 모색하는 학문인 의학이 해당되는데 '수의학'은 인간이 아닌 동물에 관한 것이므로 FD\_SCIENCE에 속한다.
  - 인의학과 수의학에는 아직 좁혀지지 않는 격차가 존재한다. → 인의학과 <수의학: FD\_SCIENCE>에는 아직 좁혀지지 않는 격차가 존재한다.
  - '인의학'은 FD\_MEDICINE으로 태깅한다.
- '생물1, 생물2'와 같이 번호가 붙은 과목의 경우 '생물'은 FD\_SCIENCE로, 번호는 QTOEDER로 각각 태깅한다.
  - 생1과 생2의 난이도 차이가 상당히 컸다. → <생: FD\_SCIENCE><1: QT\_ORDER>과 <생: FD\_SCIENCE><2: QT\_ORDER>의 난이도 차이가 상당히 컸다.
  - '생1, 생2'와 같이 줄임말로 나와도 '생물' 과목임을 알 수 있는 맥락이면 <생: FD\_SCIENCE>과 같이 줄임말 그대로 태깅한다.

## 2.2. FD\_SOCIAL\_SCIENCE

### 2.2.1. 정의 및 설명

- 사회과학 관련 학문 분야 및 학파
- 사회과학이란 인간 사회의 여러 현상을 과학적, 체계적으로 연구하는 모든 경험과학을 말한다.

### 2.2.2. 분석 대상

- 사회과학 관련 학문 분야
  - 응용사회과학: 경영학, 국제경영학, 행정학, 문헌정보학
  - 인류학, 사회인류학, 문화인류학, 사회학, 경제학, 정치학, 심리학, 지리학, 인문지리학, 지역학, 언론학, 가정학, 교육학, 법학, 세무학, 무역학, 회계학
- 사회과학 관련 학파
  - 고전학파, 신고전학파, 케임브리지학파, 뒤르켐학파, 기능학파

### 2.2.3. 예문

- 사회과학 관련 학문 분야
  - <경제학: FD\_SOCIAL\_SCIENCE>은 이와 같은 실천적 측면을 가지고 있는 것이다.
  - <통계학: FD\_SOCIAL\_SCIENCE>은 19세기 후반부터 20세기 초에 걸쳐
  - 특수교사인 주다 씨는 일반적인 <교육학: FD\_SOCIAL\_SCIENCE> 학위 외에 <특수교육

학: FD\_SOCIAL\_SCIENCE> 학위까지 가지고 있습니다.

- 사회과학 관련 학파
  - A.마셜을 창시자로 A.C.피구, J.M.케인스, D.H.로버트슨, J.V.로빈슨 등을 후계자로 하여 케임브리지대학교를 중심으로 발달하였기 때문에 <케임브리지학파: FD\_SOCIAL\_SCIENCE>라고도 한다. <케임브리지학파: FD\_SOCIAL\_SCIENCE>가 <신고전학파: FD\_SOCIAL\_SCIENCE>로 불리는 것은, 기본적으로는 이 학파가 <고전학파: FD\_SOCIAL\_SCIENCE>의 해체나 한계혁명(限界革命) 후의 영국의 정통파였으며, 한계효용 가치론을 채용하면서도 다른 학파에 비하여 공급면의 분석이나 장기동태면의 관심 등에서 <고전학파: FD\_SOCIAL\_SCIENCE>의 전통을 가장 많이 승계하고 있기 때문이다.
  - 오늘날 <사회인류학: FD\_SOCIAL\_SCIENCE>이라는 학문 분야는 <기능학파: FD\_SOCIAL\_SCIENCE>의 영향을 받은 것이다.
  - 프랑스의 <뒤르캄학파: FD\_SOCIAL\_SCIENCE>는 집단표상(集團表象)을 중시하므로 경제 현상을 사회학적 방법으로 파악하였고,

## 2.2.4. 주의 사항

- 학파는 STUDY\_FIELD로 태깅하며 이론, 사조, ~주의는 THEORY로 태깅한다. 그러나 '학파+이론'과 같이 나올 경우, 최장 단위 태깅 원칙에 따라 '학파+이론' 전체를 THEORY로 태깅한다.
- 신고전학파 성장론의 경우에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A. 새뮤얼슨이 주장하여 영향을 미친 신고전파종합(neoclassical synthesis)의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다. → <신고전학파 성장론: TR\_SOCIAL\_SCIENCE>의 경우에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 <신고전파종합: TR\_SOCIAL\_SCIENCE> (<neoclassical synthesis: TR\_SOCIAL\_SCIENCE>)의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.
  - <신고전학파: FD\_SOCIAL\_SCIENCE> <성장론: TR\_SOCIAL\_SCIENCE>으로 태깅하지 않는다.

## 2.3. FD\_MEDICINE

### 2.3.1. 정의 및 설명

- 의학 관련 학문 분야 및 학파
- 의학이란 인간을 질병으로부터 구하고 건강법을 모색하는 학문이다.
- 의학 학문분야는 양의학과 한의학을 포함하여 각 학문별 세부 학문 명칭을 나타낸다.

### 2.3.2. 분석 대상

- 학문 분야
  - 의학, 병리학, 보건의료학, 신경과학, 가정의학, 감각생리학, 기초의학, 생리학, 해부학, 생리학, 기생충학, 면역학, 미생물학, 병리학, 생화학, 약리학, 의료정보학, 의사학, 법의학, 치의학, 약학, 한의학, 기상의학, 구강위생학
  - 보건학, 약학, 내분비학, 병리학, 열대의학
- 분과
  - 비뇨기과, 신경과, 가정의학과, 간호학과, 내과, 소아청소년과, 정신과, 피부과, 재활의학과, 안과, 이비인후과, 외과, 산부인과, 영상의학과, 핵의학과, 방사선종양학과, 치과
  - ~외과(정형외과, 흉부외과 등)

### 2.3.3. 예문

- 학문 분야
  - 오늘날과 같이 <구강위생학: FD\_MEDICINE>으로서 자리를 굳히기 시작한 것은 제2차 세계대전 후부터의 일이다.
  - <보건학: FD\_MEDICINE>은 환경위생, 전염병 관리, 개인위생 교육, 질병의 조기진단과 예방을 위한 의료서비스조직,
  - <약학: FD\_MEDICINE>은 약을 중심으로 하여 화학 · 물리학 · 생물학 등의 분야를 포함하며, 그 연구 수단에 따라 학문적 체계를 구성하고 있다.
  - 오늘날까지 이 개념이 <내분비학: FD\_MEDICINE>의 기초가 되고 있다.
  - <병리학: FD\_MEDICINE>에서는 현실적으로 <병리해부학: FD\_MEDICINE>과 <실험병리학: FD\_MEDICINE>에 중점을 두고 있다.
  - 열대병을 대상으로 하는 <의학: FD\_MEDICINE>이 바로 <열대 의학: FD\_MEDICINE>이라고 하겠다.
- 분과
  - 또한 박은애 교수(1964년생, <소아청소년과: FD\_MEDICINE>)를 이대목동병원 의무부장에, 박미혜 교수(1965년생, <산부인과: FD\_MEDICINE>)를 이대목동병원 교육연구부장에 각각 임명했다.
  - <정신과: FD\_MEDICINE> 의사들이 오랫동안 믿어왔던 것 중의 하나가 '생각이 바뀌면 감정이 바뀌고 감정이 바뀌면 몸이 튼튼해진다'인데, 요즘 조금 반대입니다.
  - 일반 <치과: FD\_MEDICINE> 비용의 절반도 안 되는 파격적인 가격을 미끼로 할머니들을 현혹하는 시술업자가 포착됐다.
  - 그런데 <피부과: FD\_MEDICINE> 질환의 75%가 스트레스나 화하고 아주 밀접한 관련이 있다는 거예요.

#### 2.3.4. 주의 사항

- 신체의 일부에 해당하는 '비뇨기, 피부' 등은 FD\_MEDICINE에 해당하지 않는다. 그러나 '비뇨기와, 피부과'와 같은 의료 분과는 FD\_MEDICINE으로 태깅한다.
  - 비뇨기과는 비뇨 기관에 관한 병을 연구하는 분야다. → <비뇨기과: FD\_MEDICINE>는 비뇨 기관에 관한 병을 연구하는 분야다.
  - 피부과는 피부에 관한 모든 병을 연구하는 분야다. → <피부과: FD\_MEDICINE>는 피부에 관한 모든 병을 연구하는 분야다.
- 특정 기관의 세부 조직, 특히 교육 기관의 세부 조직으로서의 '-학과'는 태깅 대상이 아니다. 그러나 의료 분과의 경우에는 FD\_MEDICINE으로 태깅한다.
  - 그렇다면 정형외과와 재활의학과를 체험하고 → 그렇다면 <정형외과: FD\_MEDICINE>와 <재활의학과: FD\_MEDICINE>를 체험하고
- 의료 분과의 경우 FD\_MEDICINE으로 태깅하고, 구체적인 병원명인 경우 OGG\_MEDICINE으로 태깅한다.
  - 췌장정형외과에서 다리 수술을 했다. → <췌장정형외과: OGG\_MEDICINE>에서 다리 수술을 했다.

### 2.4. FD\_ART

#### 2.4.1. 정의 및 설명

- 예술 관련 학문 분야 및 학파(유파)
- FD\_ART의 하위 유형
  - 예술 관련 학문 분야
  - 예술 관련 학파(유파)

#### 2.4.2. 분석 대상

- 예술 관련 학문 분야
  - 문예학, 음악사학, 미술사학, 공연예술학
- 예술 관련 학파(유파)
  - 인상파, 점묘파, 우키요에 화파, 신경향파

#### 2.4.3. 예문

- 예술 관련 학문 분야
  - 학부 때 <공연예술: FD\_ART>을 전공하여 현재는 음악 방송의 PD가 되었다.
- 예술 관련 학파(유파)
  - 일본의 춘화는 <우키요에 화파: FD\_ART>의 등장과 더불어 활발하게 제작되었는데



- <신경향파: FD\_ART>에 드는 주요 작품을 들면 다음과 같다.

#### 2.4.4. 주의 사항

- 맥락을 고려하여 보통명사로 쓰인 것은 태깅하지 않으며 학문 분야로 쓰인 것만 FD\_ART로 태깅한다. 주로 '~을 전공하다, ~을 공부하다' 등의 술어와 같이 쓰일 경우, FD에 해당한다. 예술 분야의 경우 CV\_ART로 태깅한다.
  - 디자인을 전공하려고 미술학원에 다니고 있다. → <디자인: FD\_ART>을 전공하려고 미술학원에 다니고 있다.
  - 가격보다 디자인이 예쁜 상품이 좋다.
    - 보통명사로서 '디자인'이 쓰였기 때문에 태깅하지 않는다.
  - 한국의 고전소설은 근대 이전의 소설을 말하는 것으로 → 한국의 <고전소설: CV\_ART>은 근대 이전의 소설을 말하는 것으로
    - 예술의 한 분야로서 '고전소설'이 쓰였기 때문에 CV\_ART로 태깅한다.
- 예술의 장르가 아닌 예술 활동의 '결과물(작품)'은 FD\_ART로 태깅하지 않는다. 맥락상 '결과물'의 의미로 쓰인 보통명사는 태깅하지 않으며, 결과물의 이름은 AFW로 태깅한다.
  - 사진 700장을 발견했다. 김영하 작가의 신작 소설
    - 맥락상 '사진, 소설'은 예술 활동의 결과물이기 때문에 태깅하지 않는다.
- '학파+예술 분야'의 경우 최장 단위 태깅 원칙에 따라 '학파+예술 분야' 전체를 FD\_ART로 태깅한다.
  - 고전파음악은 균형과 조화의 형식미와 인간 내면의 객관적 표현을 중시하였다. → <고전파음악: FD\_ART>은 균형과 조화의 형식미와 인간 내면의 객관적 표현을 중시하였다.

## 2.5. FD\_HUMANITIES

### 2.5.1. 정의 및 설명

- 인간의 삶, 사고 등 인간의 근원 문제에 관해 탐구하는 학문이나 학파
- FD\_HUMANITIES의 하위 유형
  - 언어학: 국어국문학, 불어불문학, 영어영문학, 통역번역학 등
  - 역사학: 사학, 고고학, 한국사, 동양사, 서양사 등
  - 철학: 논리학, 동양철학, 서양철학, 윤리학 등
  - 종교학: 종교학, 기독교학, 불교학, 신학

### 2.5.2. 분석 대상

- 언어학 관련 학문 분야 및 학파
- 언어학, 국어국문학, 불어불문학, 영어영문학, 통역번역학

- 구조주의 언어학파, 프라하학파
- 역사학 관련 학문 분야 및 학파
- 역사학, 사학, 고고학, 한국사, 동양사, 서양사
  - 역사주의 학파, 아날학파
  - 철학 관련 학문 분야 및 학파
  - 교육철학, 형이상학, 윤리학, 정치철학, 과학철학, 언어철학, 논리학, 예술철학, 미학, 윤리학, 사회철학, 스킨라철학, 주기철학, 과학철학, 성리학
  - 스킨라학파, 플라톤학파, 신프로이트파, 유가, 도가, 묵가, 법가, 명가, 음양가, 중형가, 병가, 기호학파, 영남학파
- 종교학 관련 학문 분야 및 학파
  - 종교학, 기독교학, 카톨릭학, 불교학, 신학

### 2.5.3. 예문

- 언어학 관련 학문 분야 및 학파
- 대장금을 번역한 대만인 동문군 씨, <국어국문학: FD\_HUMANITIES> 박사학위 취득.
- <인문학: FD\_HUMANITIES>의 위기 이후로 <독어독문학: FD\_HUMANITIES>, <불어불문학: FD\_HUMANITIES>에 대한 관심이 줄어들고 있다.
- 역사학 관련 학문 분야 및 학파
  - 20여 년간 <한국사: FD\_HUMANITIES>를 가르쳐 왔습니다.
  - 여러분은 <고고학: FD\_HUMANITIES>이라는 학문을 어떻게 생각하시나요?
- 철학 관련 학문 분야 및 학파
  - <윤리학: FD\_HUMANITIES>은 일반적으로 인간의 행위에 관한 여러 가지 문제와 규범을 연구하는 학문이라고 할 수 있다.
  - <아테네학파: FD\_HUMANITIES>의 최후를 장식하는 <플라톤학파: FD\_HUMANITIES>는 의식적으로 엄밀한 정의·공준(公準)·공리에서 출발하여 논증을 하는 해석법을 연구하여 그 후의 수학 발전에 크게 공헌하였다.
  - <신프로이트파: FD\_HUMANITIES>의 일반적인 특징은 리비도를 생물적·성적으로 보는 프로이트의 고정설을 거부하고
- 종교학 관련 학문 분야 및 학파
  - 종교가 없으면서도 <종교학: FD\_HUMANITIES>을 공부하는 이유는 따로 있다.
  - <신학: FD\_HUMANITIES>을 공부하다 보면 진정한 신앙심에 대해 고민해 보게 된다.

### 2.5.4. 주의 사항

- '유교'도 FD\_HUMANITIES로 태깅한다. '기독교, 천주교, 불교' 등 다른 종교는 OGG\_RELIGION으로 태깅한다. '기독교학, 불교학'과 같이 학문일 경우에는 FD\_HUMANITIES로 태깅한다.

- '기호학파, 영남학파'는 FD\_HUMANITIES로 태깅하나 '노론, 소론'과 같은 조선시대 정치 집단은 OGG\_POLITICS로 태깅한다.
- '빈 학단'과 같은 철학 단체는 FD\_HUMANITIES의 대상이 아니다.
- '~주의'는 THEORY에 해당하지만, '주의+학파'의 경우, 최장 단위 태깅 원칙에 따라 '주의+학파' 전체를 FD\_HUMANITIES로 태깅한다.
  - 구조주의 학파 → <구조주의 학파: FD\_HUMANITIES>
- '영어, 국어'와 같은 언어의 경우, '-을 공부하다, -을 연구하다, -을 분석하다'와 같은 술어와 함께 쓰이거나 '수능, 과목, 난이도'와 같은 단어와 어울려 쓰일 경우에는 FD\_HUMANITIES로 태깅하고, '-로 말하다, -로 이야기하다'와 같은 술어와 같이 쓰일 경우에는 CV\_LANGUAGE로 태깅한다.
  - 영어가 A형이랑 B형이랑 합쳐진다고? → <영어: FD\_HUMANITIES>가 A형이랑 B형이랑 합쳐진다고?
    - 문장 내에 위에서 제시한 단어들이 나오지 않아도 문맥상 '과목'을 말하는 것이므로, FD\_HUMANITIES로 태깅한다.
  - 제니는 영어로 유창하게 말한다. → 제니는 <영어: CV\_LANGUAGE>로 유창하게 말한다.

## 2.6. FD\_OTHERS

### 2.6.1. 정의 및 설명

- 기타 학문 분야 및 학파
- FD\_SCIENCE, FD\_SOCIAL\_SCIENCE, FD\_MEDICINE, FD\_ART, FD\_HUMANITIES의 세부 학문 분야 이외의 학문 분야 및 학파

### 2.6.2. 분석 대상

- 기타 학문 분야 및 학파
- 농학, 군사학, 가정학, 건강과학, 교통학, 도서관학, 박물관학, 산림학, 스포츠과학

### 2.6.3. 예문

- 기타 학문 분야 및 학파
  - <스포츠 과학: FD\_OTHERS>은 인문, 사회, 자연과학 분야를 포괄하는 응용학문이다.
  - 이 프로그램은 <도서관학: FD\_OTHERS>, <박물관학: FD\_OTHERS>을 전공하는 대학원생을 대상으로 만들어졌다.

### 3. THEORY

---

THEORY는 특정 이론, 법칙, 원리 등의 명칭을 의미한다. 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\_MEDICINE, TR\_ART, TR\_HUMANITIES로 분류된다. 이 세부 분류를 벗어나는 이론, 법칙의 명칭은 TR\_OTHERS에 포함된다. 대학 전공 서적과 같은 전문적인 이론, 법칙 관련 용어들을 정의 또는 설명하고 있는 경우 TR로 태깅한다.

- 예제
  - 점진주의, 피마준법, 존왕양이 사상, 봉건적 배외사상, 존왕론, 양이론, 천왕숭배 사상
- 주의 사항
  - THEORY는 '~ 사상, ~ 원리, ~ 이론, ~주의'의 형태로 기술될 수 있다. 이 경우 '~ 사상, ~ 원리, ~ 이론, ~주의'를 포함해서 THEORY로 태깅한다.
  - THEORY는 '천황숭배' 사상의 형태로 기호와 함께 쓰일 수 있다. 이 경우 기호를 포함해서 THEORY로 태깅한다.

#### 3.1. TR\_SCIENCE

##### 3.1.1. 정의 및 설명

- 순수과학: 과학 관련 이론/법칙/방식/양식/체계/학설
- 기술과학: 기계/장치/시스템/기술 관련 이론/법칙/방식
- 특정 인물이 발견한 고유한 현상으로, 인물의 이름이 붙은 현상/작용(예: 도플러 효과)
  - 단순한 작용/현상은 TR\_SCIENCE의 대상으로 보지 않음.
    - 단순 작용/현상으로 TR\_SCIENCE의 대상이 아닌 예제
      - 형광작용, 중력, 빙식작용, 관성력
      - 사진작용, 이온화작용
      - 파쇄작용, 연마작용

##### 3.1.2. 분석 대상

- 순수과학
  - 특수상대성이론, 우열의 법칙, 양자론, 멘델법칙, 드무아브르의 정리, 상대성원리
  - 21세기형 수학공식, 이진법, 3원질설
- 기술과학

- 원심분리법, 기체확산법, 열확산법, 컨베이어 라인(conveyer line) 시스템
- 이온주입법, 공랭식, 수랭식

### 3.1.3. 예문

- 서로 등속도로 운동하는 관측자에 대하여 <전자기파의 이론: TR\_SCIENCE>을 포함한 모든 물리 법칙이 같은 형식으로 기술되도록 정식화되어 있다.
- 현재는 양자역학과 더불어 현대물리학, 특히 소립자물리학이나 <우주론의 지도적 원리: TR\_SCIENCE>로 간주되고 있다.
- 아인슈타인은 이 <특수상대성 원리: TR\_SCIENCE>를 발전시켜 1916년 일반상대성이론을 발표하였다.
- 운동량이나 좌표를 연산자로서 생각하는 <양자론: TR\_SCIENCE>과 대비된다.
- 동의유전자도 <멘델법칙: TR\_SCIENCE>의 변형으로 설명할 수 있다.
- <특수상대성 원리: TR\_SCIENCE>는 운동에 관한 갈릴레이-뉴턴의 <상대성 원리: TR\_SCIENCE>를 근본적으로 개혁하여, 서로 등속도로 운동하는 관측자에 대하여 전자기파의 이론을 포함한 모든 물리법칙이 같은 형식으로 기술되도록 정식화(定式化)되어 있다.
- 농축 우라늄의 제조법으로는 <기체확산법: TR\_SCIENCE>·<열확산법: TR\_SCIENCE>·<원심분리법: TR\_SCIENCE> 등의 여러 방법이 이용된다.
- 1913년 포드자동차에서 처음 적용한 <컨베이어 라인(conveyer line) 시스템: TR\_SCIENCE>이 최초이다.
- 일반적으로 행하는 <진공건조: TR\_SCIENCE>는 총체적으로 이와 같은 현상을 이용하고 있으며, 원료 중의 수분이 많은 상태에서 비교적 저온으로 건조시킬 수 있다.
- <이온주입법: TR\_SCIENCE>은 이 불순물반도체를 만드는 방법 중 하나이다.

### 3.1.4. 주의 사항

- '2진수', '자연수', '약수' 등은 태깅 대상이 아니다.
- 고유의 명칭을 갖고 있는 특정 과학 이론만이 태깅 대상이며, 여러 구체적인 이론을 포함할 수 있는 포괄적인 개념의 이론/원리는 태깅 대상이 아니다.
  - 예) 형광작용, 중력, 빙식작용, 관성력
- 특정 용어나 개념에 해당되는 것은 TR\_SCIENCE의 태깅 대상이 아니다.
- 예) 자기장, 적외선

## 3.2. TR\_SOCIAL\_SCIENCE

### 3.2.1. 정의 및 설명

- 사회과학 이론/법칙/방법/원리/사상, 정치사상

### 3.2.2. 분석 대상

- 민주주의, 공산주의, 다원적 민주주의, 법치주의, 평등주의, 관료주의, 민족주의, 자본주의, 사회주의, 개인주의, 자유주의
- 무작위추출법, 노동가치설, 대화법
- 하인리히의 법칙, 수요와 공급의 법칙
- 삼포식 농업

### 3.2.3. 예문

- <방어기제: TR\_SOCIAL\_SCIENCE>란 개인을 불안으로부터 보호하고 내적인 또는 외적인 위험이나 스트레스 요인으로부터 보호하는 자동적인 심리 과정을 뜻한다. 1단계인 <병리적인 방어기제: TR\_SOCIAL\_SCIENCE>, 2단계인 <미성숙한 방어기제: TR\_SOCIAL\_SCIENCE>, 3단계인 <신경증적 방어기제: TR\_SOCIAL\_SCIENCE>
- 개발된 시장경제와 <다원적 민주주의: TR\_SOCIAL\_SCIENCE>라는 가치관을 공유하는 국가 간 경제사회 정책 협의체로서
- 그러나 실제로는 많은 문제가 그대로 남아 있어 미국 <민주주의: TR\_SOCIAL\_SCIENCE>의 문제점이 되고 있다.
- 오늘날 <공산주의: TR\_SOCIAL\_SCIENCE>라고 할 때는 문헌에만 남아 있는 죽은 공산주의가 아니라,
- 이 <무작위추출법: TR\_SOCIAL\_SCIENCE>을 이용하여 조사대상 전체를 조사하는 대신,
- <노동가치설: TR\_SOCIAL\_SCIENCE>은 공산주의자들에 의해 지지(支持)되고 있다.
- <대화법: TR\_SOCIAL\_SCIENCE>의 원류는 고대 소크라테스의 교육 방법에서 찾을 수 있으며,
- <사회주의: TR\_SOCIAL\_SCIENCE>란 말이 <개인주의: TR\_SOCIAL\_SCIENCE>의 반대말로서 르루에 의하여 만들어졌음을 알 수 있다.

### 3.2.4. 주의 사항

- (이론) 특정 학문 분야나 이론/법칙과 같은 전문용어의 경우 그 전체가 하나의 대상을 지시하므로 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
  - 신고전학파 성장론에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 신고전파종합 (neoclassical synthesis)의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.
  - →<신고전학파 성장론: TR\_SOCIAL\_SCIENCE>에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 <신고전파종합: TR\_SOCIAL\_SCIENCE> (<neoclassical synthesis: TR\_SOCIAL\_SCIENCE>) 의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.

- (정치사상) '사회주의', '공산주의' 등의 정치사상도 TR\_SOCIAL\_SCIENCE에 해당된다.
  - 사회주의 사상에 물들어서 → <사회주의: TR\_SOCIAL\_SCIENCE> 사상에 물들어서
- (주의) '개인주의'도 자유주의 정치사상의 하위범주이므로 TR\_SOCIAL\_SCIENCE에 해당된다.
  - 개인주의의 팽배로 → <개인주의: TR\_SOCIAL\_SCIENCE>의 팽배로
- (콤플렉스) '로리타 콤플렉스', '오이디푸스 콤플렉스' 등의 콤플렉스는 심리학의 이론이므로 TR\_SOCIAL\_SCIENCE에 해당된다.
  - 그는 자신이 왜 오이디푸스 콤플렉스에 시달리는지 알지 못했다. → 그는 자신이 왜 <오이디푸스 콤플렉스: TR\_SOCIAL\_SCIENCE>에 시달리는지 알지 못했다.

### 3.3. TR\_MEDICINE

#### 3.3.1. 정의 및 설명

- 의학 요법/처방, 의학 진단법

#### 3.3.2. 분석 대상

- 팔성혈 자세교정법, 인공호흡, 요혈자극법, 뜸, 바이오피드백
- 발열요법, 견인요법, 보법, 대기요법, 공기요법
- 수지력 테스트, X 선검사
- 맥진, 내시경검사, 혈액검사, 검안, 문진

#### 3.3.3 예문

- <수지침: TR\_MEDICINE>의 방법으로는 기초기능조절 <상응요법: TR\_MEDICINE> · <기맥요법: TR\_MEDICINE> · <보사법: TR\_MEDICINE> · <오치처방: TR\_MEDICINE> · <삼일체질진단: TR\_MEDICINE> · <기맥연결요법: TR\_MEDICINE> · <팔성혈 자세교정법: TR\_MEDICINE> 등이 있다.
- 345개의 치료점에서 중요한 것은 기맥의 유주를 따른 <보사법: TR\_MEDICINE>이 있고 <요혈자극법: TR\_MEDICINE>이 있다.
- 이 <발열요법: TR\_MEDICINE>은 매독 스피로헤타가 고열에 약하다는 사실에 착안하여 발상된 것으로
- 간단한 <견인요법: TR\_MEDICINE>으로는 반창고의 응용이 있으며,
- <보법: TR\_MEDICINE>의 운용은 주로 체질과 기능의 부족을 보충하여 일체의 쇠약현상을 제거하는 법이다.
- <대기요법: TR\_MEDICINE>은 결핵환자가 자연환경에서 맑은 대기를 가까이하는 생활을 통하여 치료하는 방법이고, <공기요법: TR\_MEDICINE>은 일정한 시간에 피부를 드러내어 공기를 접하게 하는 방법이다.

- 이외에 <염파요법: TR\_MEDICINE>, <수지력 테스트: TR\_MEDICINE>에 의한 진단과 치료 및 전자진찰도 있으나 더욱 중요한 것은 음양맥진이다.
- 측각으로 환자의 질병을 판단하는 방법인 <절진: TR\_MEDICINE> 중 <맥진: TR\_MEDICINE>을 가장 중요시하기 때문에 <절진: TR\_MEDICINE>은 곧 <맥진: TR\_MEDICINE>을 말한다.
- <내시경검사: TR\_MEDICINE> 영역은 주로 기관지·식도·위·복강·흉강·요도·방광·대장·직장·자궁·관절 등이며,
- <혈액검사: TR\_MEDICINE>도 이 범주에 속하는 중요한 검사이다.
- 이 중에서 <검안: TR\_MEDICINE>의 주체가 되는 것은 굴절도의 검사인데,
- <문진: TR\_MEDICINE>의 내용은 형식적으로는 기왕력·가족병력·주소·현병력 등으로 나눈다.

### 3.4. TR\_ART

#### 3.4.1. 정의 및 설명

- 예술 관련 이론/법칙/방식/양식, 예술사조, 촬영/연출 기법

#### 3.4.2. 분석 대상

- 에로티시즘, 원근법, 대기원근법, 선원근법, 구름법, 물골법
- 대위법, 데카당스, 문자기보법, 환유법, 제유법
- 고딕양식, 고전주의, 유량악보, 낭만주의
- 몽타주, 오마주, 추사체, 데칼코마니, 액션페인팅
- 익스트림 롱 샷, 풀 피겨 샷, 클로즈업 샷, 루즈샷, 팬닝, 틸트

#### 3.4.3 예문

- <고딕양식: TR\_ART>의 발달로 교회의 외관은 특색이 강해졌고
- <고전주의: TR\_ART>의 부활이야말로
- <원근법: TR\_ART>은 넓은 의미로 볼 때 화면 깊이의 표현이며, <대기원근법: TR\_ART>은 <선원근법: TR\_ART>의 보조적 효과를 가질 뿐만 아니라, 화면에 중심과 통일을 주어 고전 양식을 성립하게 한다.
- 영화에서 <에로티시즘: TR\_ART>은 존재의 문제에 접근하려는 시도를 보였으나 때로는 성 행위를 노골적으로 묘사한 포르노그래피에 불과한 경우도 많았다.
- <대위법: TR\_ART>에서는 각 성부가 명료하게 식별할 수 있는 선율적 독립성을 지니며,
- 이후 그들의 예술적 경향을 <데카당스: TR\_ART>라고 평하였다.
- 이 <문자기보법: TR\_ART>에 의한 악보를 문자보라고 한다.
- <환유법: TR\_ART>은 표현하고자 하는 사물의 특징으로 전체를 나타내는 방법이지만(예:



금수강산→대한민국, 요람→탄생, 무덤→죽음), 제유법은 같은 종류의 사물 중에서 어느 한 부분을 들어 전체를 나타내는 방법이다.

- 예술의 장에서 브레이크가 이것을 <소외효과: TR\_ART>라고 이론화한다.

#### 3.4.4. 주의 사항

- (-이즘) '다다이즘' 등의 예술운동은 일종의 예술사조로 고려해서 TR\_ART로 태깅한다.
  - 다다이즘의 예술운동은 잘 알려지지 않았지만 → <다다이즘: TR\_ART>의 예술운동은 잘 알려지지 않았지만
- (사회운동, 변혁) '르네상스'는 예술운동 외의 학문, 사상 등까지 다루고 있으므로 예술사조로 보기 힘들다. '르네상스'는 일종의 문화혁신운동이므로 'CV\_CULTURE'로 태깅한다
  - <르네상스: CV\_CULTURE>는 14세기 중반에 인문주의를 바탕으로

### 3.5. TR\_HUMANITIES

#### 3.5.1. 정의 및 설명

- 철학, 언어, 역사, 종교적 이론/사상

#### 3.5.2. 분석 대상

- 조작주의, 주지주의, 구조주의, 토마스주의, 경험주의, 합리주의, 공리주의, 현실주의
- 니힐리즘, 이데아설, 인식론, 노장사상, 이기론, 존재론, 관념론

#### 3.5.3 예문

- 엄밀한 의미에서의 <니힐리즘: TR\_HUMANITIES>은 아무것도 존재하지 않는다는, 즉 '무'라는 주장이다.
- <조작주의: TR\_HUMANITIES>는 원래 물리학의 상대성이론에서 여러 개념을 규정하는 데 사용되었으나,
- <이데아설: TR\_HUMANITIES>을 소크라테스의 것이라고 하여 버넷-테일러설을 세웠다.
- J.로크의 《인간오성론》은 이 인식문제의 전환점에 선 것이었고, D.흄 등의 영국 경험론에 의하여 <인식론: TR\_HUMANITIES>의 근대적 성격은 더욱 명확하게 되었으며 I.칸트에서 대성되었다.
- 그러나 그들의 사상이 도가사상을 실질적으로 대표하므로 <노장사상: TR\_HUMANITIES>을 일반적으로 도가사상과 같은 것으로 보기도 한다.
- 이에 따라 이와 기를 유기적으로 결합한 <이기론: TR\_HUMANITIES>이 확립되었다.
- 인식이 감관에 의한 것이 아니라 지성에 의해서 생긴다고 보는 합리론도 넓은 뜻의 <주지주의: TR\_HUMANITIES>이다.

#### 3.5.4. 주의 사항

- (학파, 철학 유파) “플라톤 학파” 등의 철학 유파는 FD\_HUMANITIES로 태깅한다.
  - 플라톤 학파의 영향을 받아 탄생한 → <플라톤 학파: FD\_HUMANITIES>의 영향을 받아 탄생한
- (학파) ‘유가, 도가, 묵가, 법가, 명가, 음양가, 종횡가, 병가’ 역시 학파이기 때문에 FD로 태깅한다.
  - 도가가 영향을 끼쳐 탄생하였다. → <도가: FD\_HUMANITIES>가 영향을 끼쳐 탄생하였다.

### 3.6. TR\_OTHERS

#### 3.6.1. 정의 및 설명

- 기타 이론/법칙/원리
- TR\_SCIENCE, TR\_SOCIAL\_SCIENCE, TR\_MEDICINE, TR\_ART, TR\_HUMANITIES의 세부 이론 분야 이외의 이론/법칙/원리

#### 3.6.2. 분석 대상

- 샤머니즘 등

#### 3.6.3. 예문

- 토착종교인 <샤머니즘: TR\_OTHERS>의 고원교를 몰아내는 투쟁과 그 승리를 극명하게 표현한다.

#### 3.6.4. 주의 사항

- TR\_SOCIAL\_SCIENCE나 TR\_HUMANITIES로 태깅하지 않는다.
  - (샤머니즘) 토착종교인 샤머니즘의 고원교를 몰아내는 투쟁과 그 승리를 극명하게 표현한다. → 토착종교인 <샤머니즘: TR\_OTHERS>의 고원교를 몰아내는 투쟁과 그 승리를 극명하게 표현한다.

## 4. ARTIFACTS

---

ARTIFACTS는 사람에 의해 창조된 인공물을 나타낸다. 문화재, 건물, 악기, 도로, 무기, 운송수단, 작품명, 공산품명 등이 이에 해당된다. ARTIFACTS 세부범주에 해당되지 않는 인공물은 태깅 대상에서 배제한다.

### 4.1. AF\_BUILDING

#### 4.1.1. 정의 및 설명

- 건축물/토목건설물
- 건축물/토목건설물의 하위 유형
  - 건물
  - 다리
  - 댐
  - 기타 건설물

#### 4.1.2. 분석 대상

- 건물 이름
  - 한화생명 63빌딩, 롯데월드타워, Two IFC, 서울종합운동장 야구장, 잠실실내체육관, 대우 트림프월드 I, 은마아파트, 백악관, 파리 노트르담 대성당 등
- 다리 이름
  - 성수대교, 한강철교, 폰네프, 골든 게이트 브리지 등
- 댐 이름
  - 소양강댐, 팔당댐, 평화의댐, 상주보, 후버댐, 상주보, 공주보, 세종보 등
- 기타 건설물
  - 워싱턴 기념탑, 베를린 장벽, 잠원 한강공원 수영장 등

#### 4.1.3. 예문

- 건물 이름
  - 샷포로의 주소 체계는 <샷포로 TV 타워: AF\_BUILDING>를 기준으로 이루어졌다.
  - <백악관: AF\_BUILDING> 앞은 시위 행렬로 가득했다.
  - 1995년 6월 29일에 <삼풍백화점: AF\_BUILDING>이 무너졌다.
  - 대치동 <은마아파트: AF\_BUILDING>는 당시 물가의 비교 기준으로 자주 거론된다.
- 다리 이름
  - 할리우드 영화에는 <골든게이트 브리지: AF\_BUILDING>가 자주 등장한다.

- <성수대교: AF\_BUILDING> 붕괴 사고로 등교 중이던 무학여고 학생들이 다수 사망했다.
- 댐 이름
  - <소양강댐: AF\_BUILDING>은 대한민국에서 가장 큰 사력식 댐이다.
  - 전두환 정권은 <평화의 댐: AF\_BUILDING>을 짓겠다며 아이들의 코 묻은 돈까지 빼앗았다.
  - <합천창녕보: AF\_BUILDING>는 물의 흐름을 방해하여 큰 홍수를 유발했다.
  - 기타 건설물
  - <워싱턴 기념탑: AF\_BUILDING>은 리플렉팅 풀을 사이에 두고 링컨기념관과 서로 마주보고 있다.
  - 베를린 시의 기증으로 <베를린 장벽: AF\_BUILDING>의 일부가 서울 중심부에 전시되어 있다.

#### 4.1.4. 주의 사항

- 어떠한 단어가 건축물(AF)이면서 기관(OG)의 의미를 지닐 경우, '~이 보이대|붕괴하다|흔들리다, ~을 짓다|만들다'와 함께 쓰이는 등 문맥에 의해 건축물의 의미로 해석되는 경우만 AF로 태깅한다.
  - 청와대 본관은 1991년에 신축되었다. → <청와대 본관: AF\_BUILDING>은 1991년에 신축되었다.
  - 인왕산에 오르면 동쪽으로 청와대가 보인다. → 인왕산에 오르면 동쪽으로 <청와대: AF\_BUILDING>가 보인다.
  - 그린벨트 해제, 한발 물러선 청와대 → 그린벨트 해제, 한발 물러선 <청와대: OGG\_POLITICS>
- 건축물이 지명이나 상위 조직과 함께 쓰인 경우, 전체가 정식 명칭이면 분리하지 않고 하나로 태깅한다.
  - 잠원 한강공원 수영장 → <잠원 한강공원 수영장: AF\_BUILDING>
  - 한강 잠원동 수영장 → <한강: LCG\_RIVER> <잠원동: LCP\_COUNTY> 수영장
    - '수영장'은 정식 고유 명칭이 아니므로 태깅하지 않는다.
- 일상적으로 많이 쓰이더라도 정식 명칭이 아니면 전체를 하나의 명칭으로 태깅하지 않는다.
  - 서울 올림픽공원 체조경기장 → <서울: LCP\_CAPITALCITY> <올림픽공원: LC\_OTHERS> 체조경기장

## 4.2. AF\_CULTURAL\_ASSET

### 4.2.1. 정의 및 설명

- 건축물, 고분, 사찰 등의 문화재
- 문화재의 하위 유형
  - 건축/건설물
  - 고분
  - 문화재로 지정되어 있지 않은 조선시대까지 만들어진 절과 무덤
  - 기타 해외 문화재

#### 4.2.2. 분석 대상

- 건축물
  - 건물: 송례문, 영주 부석사 무량수전, 합천 해인사 장경판전 등
  - 탑: 서울 원각사지 십층석탑, 경주 불국사 삼층석탑, 장흥 보림사 남·북 삼층석탑 및 석 등 등
  - 기타 건조물: 경주 석굴암 석굴, 경주 불국사 청운교 및 백운교, 청주 용두사지 철당간 등
- 고분
  - 무령왕릉, 천마총, 가야고분, 신장절공묘역 등
  - 문화재로 지정되어 있지 않은 조선시대 이전에 만들어진 절과 무덤
  - 상원사, 보문사, 단월동 고묘군, 거창 석강리 고분군 등
- 기타 해외문화재
  - 에펠탑, 자유의 여신상, 미디 운하 등

#### 4.2.3. 예문

- 건축물
  - 남대문 시장을 지나 <송례문: AF\_CULTURAL\_ASSET> 앞에 도착한다.
  - <경주 불국사 삼층석탑: AF\_CULTURAL\_ASSET>은 한국에서 가장 유명한 탑이라고 해도 과언이 아니다.
  - <용두사지 철당간: AF\_CULTURAL\_ASSET>은 시내 중심가에 우뚝 솟아 있는데 주변 높은 건물들 때문에 잘 보이지는 않는다.
- 고분
  - <천마총: AF\_CULTURAL\_ASSET>에서 시작하여 흥융들로 이어지는 포석로를 황리단길이라고 부른다.
- 문화재로 지정되어 있지 않은 조선시대 이전에 만들어진 절과 무덤
  - <상원사: AF\_CULTURAL\_ASSET> 문수전 앞 고양이 석상이 관광객들 사이에서 인기를 끌고 있다 .
- 기타 해외문화재
  - <자유의 여신상: AF\_CULTURAL\_ASSET>이 흰색이라고 생각하는 사람들이 많다.

#### 4.2.4. 주의 사항

- (건축물) 절이나 탑, 운하 등 건축 문화재는 AF\_CULTURAL\_ASSET으로 태깅하고 공예품 등 나머지 문화재는 다른 세부 개체명을 우선시하여 태깅한다.
  - 무량수전 앞에서 집합 → <무량수전: AF\_CULTURAL\_ASSET> 앞에서 집합
  - 진흥왕 순수비는 교과서에도 나온다. → <진흥왕 순수비: AFA\_ART\_CRAFT>는 교과서에도 나온다.
  - 봉산탈춤은 황해도 봉산군의 전통 춤이다. → <봉산탈춤: AFA\_PERFORMANCE>은 황해도 봉산군의 전통 춤이다.
- (건축물) ARTIFACTS에 해당하는 대상이 AF\_CULTURAL\_ASSET인지 AF\_BUILDING인지 문맥으로 판단할 수 없는 경우에는 본유적 의미인 AF\_BUILDING을 우선하여 태깅한다.
  - 관덕정 앞에서 찍은 가족사진 → <관덕정: AF\_BUILDING>앞에서 찍은 가족사진
    - 보물인 '제주 관덕정'과 국가등록문화재인 '광주 관덕정'은 AF\_CULTURAL\_ASSET으로 태깅해야 하고, '대구 관덕정' 외 기타 개인 시설은 문화재가 아니므로 AF\_BUILDING으로 태깅해야 한다. 예문의 '관덕정'은 다른 설명이 없어 문화재인지 판단이 불가하므로 AF\_BUILDING으로 태깅한다.
- (자연유산) 동식물 등의 자연 유산은 PLANT, ANIMAL 계열의 세부 개체명으로 태깅하고, 동식물 서식지 및 자생지는 AF\_CULTURAL\_ASSET으로 태깅한다.
  - 설악산국립공원의 깃대종인 산양은 천연기념물로 지정되어 있다. → 설악산국립공원의 깃대종인 <산양: AM\_MAMMALIA>은 천연기념물로 지정되어 있다.
  - 양평 용문사 은행나무의 나이는 1,100년으로 추정된다. → 양평 <용문사: AF\_CULTURAL\_ASSET> <은행나무: PT\_TREE>의 나이는 1,100년으로 추정된다.
  - 제주도에 온 관광객들은 만장굴과 정방폭포에 꼭 들른다. → 제주도에 온 관광객들은 <만장굴: AF\_CULTURAL\_ASSET>과 <정방폭포: AF\_CULTURAL\_ASSET>에 꼭 들른다.
  - 제주 토끼섬 문주란 자생지 → <제주 토끼섬 문주란 자생지: AF\_CULTURAL\_ASSET>
  - 정선 정암사 열목어 서식지 → <정선 정암사 열목어 서식지: AF\_CULTURAL\_ASSET>
  - 제주 연안 연산호 군락 → <제주 연안 연산호 군락: AF\_CULTURAL\_ASSET>
- (인간문화재) 인간문화재는 '인간문화재 000'과 같이 표현되므로 무형문화재를 보유한 사람 이름을 PS\_NAME으로 태깅한다.
  - 인간문화재 공옥진 → 인간문화재 <공옥진: PS\_NAME>
- 서책과 함이 함께 하나의 문화재로 지정되는 등 하나의 세분류로 특정할 수 없는 문화재는 전체를 AF\_CULTURAL\_ASSET으로 태깅한다.
  - 이형 좌명원종공신녹권 및 함 → <이형 좌명원종공신녹권 및 함: AF\_CULTURAL\_ASSET>
- '보물 1호', '국보 1호, 세계문화유산' 등에서 '보물', '국보', '세계문화유산' 등은 태깅 대상이 아니다.
  - 국보 제4호 여주 고달사지 승탑 → 국보 <제4호: QT\_ORDER> <여주 고달사지 승탑: AF\_CULTURAL\_ASSET>

## 4.3. AF\_ROAD

### 4.3.1. 정의 및 설명

- 도로, 철로, 수로 등 길의 명칭
- 길의 하위 유형
  - 도로
  - 철로
  - 수로
  - 관광지 거리/골목명

### 4.3.2. 분석 대상

- 도로
  - 고속화 도로 이름: 경부고속도로, 서울양양고속도로, 올림픽대로, 내부순환로 등
  - 국도 이름: 국도1호선, 46번국도 등
  - 도로명주소 이름: 테헤란로, 강남대로, 세종대로 등
  - 삼거리, 사거리, 교차로 등
- 철로
  - 일반 철도: 경부선, 경춘선, 장항선, 중앙선, 남부화물기지선 등
  - 도시 철도: 수도권 전철 1호선, 서울 지하철 2호선, 신분당선, 용인경전철 등
- 수로
  - 운하: 파나마 운하, 수에즈 운하, 경인운하, 한반도 대운하, 경인 아라뱃길 등
- 관광지 거리/골목명
  - 가로수길, 황리단길, 경리단길, 청량리 먹자골목, 제주 올레길, 북한산 둘레길 등

### 4.3.3. 예문

- 도로 이름
  - <3번국도: AF\_ROAD>의 주요 정체 구간은 송내삼거리에서 봉양사거리 구간이다.
  - 세종의 시호를 따서 <세종대로: AF\_ROAD>라는 명칭을 붙였다.
  - 법원과 검찰청이 떠나게 되면 <법원검찰청 사거리: AF\_ROAD>도 이름을 바꿀 수밖에 없다.
  - 국립국어원의 도로명 주소는 서울특별시 강서구 <금남화로: AF\_ROAD> 154이다.
- 철로 이름
  - <경춘선: AF\_ROAD> ITX-청춘열차를 주로 이용한다.
  - <서울 지하철 2호선: AF\_ROAD>에만 40개가 넘는 역이 있다.
- 수로 이름
  - <파나마 운하: AF\_ROAD>는 대서양과 태평양을 연결한다.

- 관광지 거리/골목명
  - 경주를 찾는 젊은이들은 석굴암, 첨성대 등 기존 관광지보다 <황리단길: AF\_ROAD>을 먼저 찾는다.

#### 4.3.4. 주의 사항

- 도로명주소법에서의 도로명과 도로명에 붙은 'n길', 'n로' 등은 AF\_ROAD로, 도로명 다음에 나오는 상세 번지와 건물 번호 및 동-호수는 QT\_ADDRESS로 태깅한다.
  - 남대문로3길 9-1 → <남대문로3길: AF\_ROAD> <9-1: QT\_ADDRESS>
  - 서울특별시 강서구 금남화로 154 101동 205호 → <서울특별시: LCP\_CAPITALCITY> <강서구: LCP\_COUNTY> <금남화로: AF\_ROAD> <154: QT\_ADDRESS> <101동: AF\_BUILDING> <205호: QT\_ADDRESS>
  - (06288) 서울시 강남구 삼성로1길 150 대치미도 1차 아파트 103동 101호 → (<06288: QT\_ADDRESS>) <서울시: LCP\_CAPITALCITY> <강남구: LCP\_COUNTY> <삼성로1길: AF\_ROAD> <150: QT\_ADDRESS> <대치미도 1차 아파트 103동: AF\_BUILDING> <101호: QT\_ADDRESS>
- '국도1호선, 44번 국도' 등 고유의 번호를 가진 국도는 통합하여 태깅한다.
  - 11번 국도 → <11번 국도: AF\_ROAD>
- 사거리, 로터리 등 길의 명칭이 맥락에 따라 '~에 있다/위치하다' 등과 함께 쓰여 특정 위치를 의미하는 경우는 LC\_OTHERS로 태깅한다.
  - 미아사거리에 있는 현대백화점에서 만나자. → <미아사거리: LC\_OTHERS>에 있는 현대백화점에서 만나자.
- 제주 해안도로, 북한산 등산로, 이기대 산책로 등 특정한 길을 지칭하지 않는 경우에는 AF\_ROAD로 태깅하지 않는다.
  - 북한산국립공원백운대코스가 북한산 등산로 중 가장 험하다. → <북한산국립공원백운대코스: AF\_ROAD>가 북한산 등산로 중 가장 험하다.
    - 북한산에는 수많은 등산로가 있으므로, '북한산 등산로'는 특정 길을 지칭하지 않는다. 따라서 '북한산 등산로'는 태깅 대상이 아니다.
- '명동거리', '덕수궁돌담길', '경리단길' 등 길에 위치한 다른 즐길 거리/시설을 포함하는 명칭도 AF\_ROAD로 태깅한다.
  - 요즘 젊은이들 사이에서 <경리단길: AF\_ROAD> 투어가 핫하다.

### 4.4. AF\_TRANSPORT

#### 4.4.1. 정의 및 설명

- 운송 수단 및 놀이기구 등 탈것



- 운송 수단의 하위 유형
  - 육상 운송 수단
  - 수상 운송 수단
  - 공중 운송 수단
- 놀이기구

#### 4.4.2. 분석 대상

- 운송 수단의 하위 유형
  - 육상 운송 수단: 자전거, 자동차, 기차, 소나타, 포니, 카니발 등
  - 수상 운송 수단: 유람선, 화물선, 전함, 세월, 광개토대왕, 을지문덕함 등
  - 공중 운송 수단: 비행기, 헬리콥터, 비행선, 콩코드 등
- 놀이기구의 하위 유형
  - 롤러코스터, 바이킹, 범퍼카 등

#### 4.4.3. 예문

- 운송 수단의 하위 유형
  - 휴갓길 교통수단으로 개인 <승용차: AF\_TRANSPORT>보다 <고속버스: AF\_TRANSPORT>나 <기차: AF\_TRANSPORT>를 선호하는 사람들이 많아지고 있다.
  - 한강에서 <유람선: AF\_TRANSPORT>을 타 보지 않은 사람이 진정한 서울 사람이라는 말도 나온다.
  - 김포공항에서는 국내선 <비행기: AF\_TRANSPORT>와 일부 국외선 <비행기: AF\_TRANSPORT>를 이용할 수 있다.
- 놀이기구의 하위 유형
  - 놀이기구가 다양해졌지만 여전히 사람들이 가장 사랑하는 놀이기구는 <롤러코스터: AF\_TRANSPORT>이다.

#### 4.4.4. 주의 사항

- 'KTX', '세월호', '카니발' 등 운송 수단 중 고유명(상품명)이 쓰인 경우에도 AF\_TRANSPORT로 태깅한다.
  - KTX는 선로 위를 300~305km/h의 속도로 달린다. → <KTX: AF\_TRANSPORT>는 선로 위를 300~305km/h의 속도로 달린다.
  - 세월호 침몰로 단원고 학생 325명이 숨졌다. → <세월호: AF\_TRANSPORT> 침몰로 단원고 학생 325명이 숨졌다.
  - 기아자동차에서 4세대 카니발을 새롭게 출시했다. → 기아자동차에서 4세대 <카니발: AF\_TRANSPORT>을 새롭게 출시했다.
- 운송수단이 아닌 인공위성은 TMI\_HW로 태깅한다.

- 과학기술 3호 → <과학기술 3호: TMI\_HW>
- 무기를 갖춘 함선도 AF\_TRANSPORT로 태깅한다.
  - 세종대왕급 구축함으로 초도함 세종대왕과 2번함 울곡이이, 3번함 서애류성룡이 취역하면서 해상기동부대의 전략적 운용이 한층 강화될 것으로 전망된다. → 세종대왕급 <구축함: AF\_TRANSPORT>으로 초도함 <세종대왕: AF\_TRANSPORT>과 2번함 <울곡이이: AF\_TRANSPORT>, 3번함 <서애류성룡: AF\_TRANSPORT>이 취역하면서 해상기동부대의 전략적 운용이 한층 강화될 것으로 전망된다.

## 4.5. AF\_MUSICAL\_INSTRUMENT

### 4.5.1. 정의 및 설명

- 악기 명칭
- 악기의 하위 유형
  - 현악기
  - 관악기
  - 타악기

### 4.5.2. 분석 대상

- 현악기
  - 바이올린, 비올라, 거문고, 피아노, 해금, 아쟁 등
  - 스트라디바리우스 레이디 블런트(바이올린명), 팬더 스트라토캐스터(기타명) 등
- 관악기
  - 플루트, 피콜로, 클라리넷, 색소폰, 대금, 단소 등
- 타악기
  - 북, 편경, 팀파니, 마림바, 실로폰, 심벌즈, 탬버린 등

### 4.5.3. 예문

- 현악기
  - <해금: AF\_MUSICAL\_INSTRUMENT>은 현악기이면서도 관악합주에 반드시 편성되어 관악기와 현악기와의 균형을 유지하며
  - <바이올린: AF\_MUSICAL\_INSTRUMENT>이 <첼로: AF\_MUSICAL\_INSTRUMENT>보다 높은 음역대를 담당한다.

#### 관악기

- 이 위에 갈대 속청을 붙여 이것으로 <대금: AF\_MUSICAL\_INSTRUMENT> 특유의 음색을 낸다.

- <오보에: AF\_MUSICAL\_INSTRUMENT>란 이름은 프랑스어의 '음이 높은 나무 피리'라는 뜻의 오브와(hautbois)에서 나왔다.
- 타악기
  - 중국의 <석경: AF\_MUSICAL\_INSTRUMENT>보다 좋은 <편경: AF\_MUSICAL\_INSTRUMENT>을 만들어 내게 되었다.
  - <괭과리: AF\_MUSICAL\_INSTRUMENT>가 종묘제례악(宗廟祭禮樂)에 쓰일 때는 다른 이름으로 불린다.

#### 4.5.4. 주의 사항

- '~을 연주하다|켜다|타다|치다|불다' 등의 서술어와 함께 쓰인다.
  - 그 노래는 배우가 드라마에서 직접 기타를 치며 노래를 불러 유명해졌다. → 그 노래는 배우가 드라마에서 직접 <기타: AF\_MUSICAL\_INSTRUMENT>를 치며 노래를 불러 유명해졌다.
- 악기명은 대체로 보통명사이지만, '스트라디바리우스 레이디 블런트(바이올린명)', '팬더 스트라토캐스터(기타명)' 등과 같이 고유명이 쓰이는 경우도 있을 수 있다. 두 경우 모두 AF\_MUSICAL\_INSTRUMENT로 태깅한다.
  - 지난 2011년 경매에 나온 '레이디 블런트'는 기존 '스트라디바리우스'의 경매가를 넘어서며 바이올린 사상 최고가를 기록했다. → 지난 2011년 경매에 나온 '<레이디 블런트: AF\_MUSICAL\_INSTRUMENT>'는 기존 '<스트라디바리우스: AF\_MUSICAL\_INSTRUMENT>'의 경매가를 넘어서며 <바이올린: AF\_MUSICAL\_INSTRUMENT> 사상 최고가를 기록했다.

### 4.6. AF\_WEAPON

#### 4.6.1. 정의 및 설명

- 무기의 명칭
- 무기의 하위 분류
  - 도검류
  - 총포류
  - 폭탄류
  - 기타 무기류

#### 4.6.2. 분석 대상

- 도검류: 칼, 검, 도끼, 창 등
- 총포류: 총, 곡사포, 고정포, K-1, K-2, M16A1, AK-47 등
- 폭탄류: 수류탄, 시한폭탄, 어뢰, 탄도미사일, 비격진천뢰 등

- 기타 무기류: 최루가스, 신경작용제, 핵무기 등

#### 4.6.3. 예문

- 도검류
  - <칼: AF\_WEAPON>에 찔리면 대부분 파상풍 등 감염병에 의해 죽는다.
- 총포류
  - 한국은 경찰도 마음대로 <총: AF\_WEAPON>을 휴대할 수 없다.
- 폭탄류
  - <탄도미사일: AF\_WEAPON>은 제2차 세계대전 중에 독일이 개발하여 런던 공격에 최초로 사용되었다.
- 기타 무기류
  - <최루가스: AF\_WEAPON>를 사용할 경우에는 가스탄 · 투척 · 투사기(投射器) 등이 사용된다.

#### 4.6.4. 주의 사항

- (무기의 고유명) 'k-2', 'M16A1'처럼 무기의 고유 명칭(상품명)이 쓰인 경우에도 AF\_WEAPON로 태깅한다.
  - 한국형 소총 K-2는 대한민국 육군의 기본 화력을 구성한다. → 한국형 <소총: AF\_WEAPON> <K-2: AF\_WEAPON>는 대한민국 육군의 기본 화력을 구성한다.
- (문화재) '승자총통', '비격진천뢰'와 같이 조선시대 이전에 제작되어 문화재로 등록된 무기의 고유 명칭도 AF\_WEAPON으로 태깅한다.
  - 1583년(선조 16) 6월 승자총통을 처음 만들었다는 기록이 있다. → 1583년(선조 16) 6월 <승자총통: AF\_WEAPON>을 처음 만들었다는 기록이 있다.
- 특정 문맥에서만 무기로 볼 수 있는 보통명사도 AF\_WEAPON으로 태깅한다.
  - 장군의 명령에 따라 농민들도 호미와 낫을 들고 왜군과 싸웠다. → 장군의 명령에 따라 농민들도 <호미: AF\_WEAPON>와 <낫: AF\_WEAPON>을 들고 왜군과 싸웠다.

### 4.7. AF\_ART\_WORKS

- 작품명은 AF\_ART\_WORKS의 하위 부류인 AFA로 태깅한다.

#### 4.7.1. AFA\_DOCUMENT

##### 4.7.1.1. 정의 및 설명

- 도서, 문학 작품, 기타 문서의 명칭

- 도서의 하위 유형
  - 철학 서적명
  - 역사 서적명
  - 사회과학 서적명
  - 자연과학 서적명
  - 공업 서적명
  - 산업 서적명
  - 예술 서적명
  - 어학 서적명
  - 문학 서적명
- 문학 작품의 하위 유형
  - 시의 제목
  - 소설의 제목
  - 희곡의 제목

#### 4.7.1.2. 분석 대상

- 도서의 하위 유형
  - 철학 서적명: 존재와 시간, 대학, 논어, 도덕경 등
  - 역사 서적명: 세종실록, 고려사 등
  - 사회과학 서적명: 코란, 한역대장경, 개역성서, 대동여지도 등
  - 자연과학 서적명: 동의보감, 본초강목 등
  - 공업 서적명: 프로그래밍의 이해, 정보 처리 입문 등
  - 산업 서적명: 산업안전기사 실기, 기계설비기술사 수험서 등
  - 예술 서적명: 미학오디세이, 신화의 미술관 등
  - 어학 서적명: 꼬리에 꼬리를 무는 영문법, 한국어 글쓰기 등
  - 문학 서적명: 채식주의자, 지금 알고 있는 걸 그때도 알았더라면 등
- 문학 작품의 하위 유형
  - 시의 제목: 꽃을 보듯 너를 본다, 님의 침묵 등
  - 소설의 제목: 데미안, 봉제인형 살인사건 등
  - 희곡의 제목: 살아있는 이중생 각하, 국물있사옵니다 등

#### 4.7.1.3. 예문

- 철학 서적명
  - 그의 심현한 철학사상서이자 우수한 문학서인 이 《<장자: AFA\_DOCUMENT>》는 위(魏)·진(晉) 때에 널리 읽히고 육조시대(六朝時代)까지 그 사상이 유행하였다.
- 사회과학 서적명

- 《<신증동국여지승람: AFA\_DOCUMENT>》에 전라도의 토산물로 기록되어 있다
- 자연과학 서적명
  - 《<자산어보: AFA\_DOCUMENT>(<茲山魚譜: AFA\_DOCUMENT>)》에 보면, 맛이 달콤하고 회·국·포를 만들기 좋다고 했으며, 《<동의보감: AFA\_DOCUMENT>(<東醫寶鑑: AFA\_DOCUMENT>)》에 성(性)이 평(平)하고 맛이 달며 독이 없다고 했다.
- 소설의 제목
  - 장편 《<난장이가 쓰아올린 작은 공: AFA\_DOCUMENT>》은 12편의 작품으로 구성된 연작소설로 도시빈민의 가난한 삶과 처참한 패배의 한을 다루고 있다.
- 희곡의 제목
  - 대본은 프랑스의 극작가 A.E.스크리브의 《<구스타브 3 세: AFA\_DOCUMENT>》를 A.소마가 이탈리아어로 번역한 것이다.

#### 4.7.1.4. 주의 사항

- 만화책, 웹툰은 AFA\_DOCUMENT로 태깅하고, 애니메이션은 AFA\_VIDEO로 태깅한다.
  - 90년대에 태어난 아이들은 드래곤볼, 슬램덩크 등 일본 만화책에 익숙하다. → 90년대에 태어난 아이들은 <드래곤볼: AFA\_DOCUMENT>, <슬램덩크: AFA\_DOCUMENT> 등 일본 만화책에 익숙하다.
  - TV 앞에는 겨울왕국을 보려는 아이들이 모여 들었다. → TV 앞에는 <겨울왕국: AFA\_VIDEO>을 보려는 아이들이 모여 들었다.
- 서적으로서의 <훈민정음>은 AFA\_DOCUMENT로 태깅하지만, 문자적인 의미의 '훈민정음'은 개체명이 아니므로 태깅하지 않는다.
  - 글자 훈민정음과 책 훈민정음은 구분할 필요가 있다. → 글자 훈민정음과 책 <훈민정음: AFA\_DOCUMENT>은 구분할 필요가 있다.
  - 한국어를 표기하기 위해 만들어진 한글의 다른 이름은 훈민정음이다. → 한국어를 표기하기 위해 만들어진 한글의 다른 이름은 훈민정음이다.
    - 이때의 '훈민정음'은 문자의 의미로 쓰였으므로 태깅하지 않는다.
- 연극 작품명이 명확한 경우에만 AFA\_PERFORMANCE로 태깅하고, 그렇지 않으면 시나리오로 보아 AFA\_DOCUMENT로 태깅한다.
  - 김 작가는 셰익스피어의 로미오와 줄리엣을 드라마화했다. → 김 작가는 셰익스피어의 <로미오와 줄리엣: AFA\_DOCUMENT>을 드라마화했다.
  - 이번 졸업작품으로 로미오와 줄리엣이 주말 동안 2회 공연된다. → 이번 졸업작품으로 <로미오와 줄리엣: AFA\_PERFORMANCE>이 주말 동안 2회 공연된다.

### 4.7.2. AFA\_PERFORMANCE

#### 4.7.2.1. 정의 및 설명

- 춤, 무용, 연극 등 공연의 작품명
- 공연의 하위 유형
  - 무용명
  - 연극명
  - 뮤지컬명
  - 기타 공연의 성격을 가지는 무형문화재

#### 4.7.2.2. 분석 대상

- 무용명: 바라춤, 동래학춤, 농악무 등
- 연극명: 고도를 기다리며, 레베카, 양주별산대놀이, 통영오광대 등
- 뮤지컬명: 오페라의 유령, 레미제라블 등
- 기타 공연의 성격을 가지는 무형문화재: 단오제, 씨김굿, 줄타기 등

#### 4.7.2.3. 예문

- 무용명
  - <파소도블레: AFA\_PERFORMANCE>는 스페인의 전통 춤의 일종이다. 남녀가 짝을 지어 8분의 6박자라는 빠른 템포에 맞춰 1분에 약 120번 스텝을 밟는다.
- 연극명
  - 연극 '<고도를 기다리며: AFA\_PERFORMANCE>'는 인간 존재의 부조리성을 보여 주는 사무엘 베케트의 대표작이다.
- 뮤지컬명
  - 1990년 내한한 영국산 뮤지컬 《<캣츠: AFA\_PERFORMANCE>》는 예술의 전당 오페라극장에서 서구 뮤지컬의 진수를 만끽하게 해 주었다.
- 기타 공연의 성격을 가지는 무형문화재
  - 한국의 전통공연예술인 <줄타기: AFA\_PERFORMANCE>는 공연자의 노래와 춤, 이야기가 한데 어우러진 종합예술의 성격을 보인다.

#### 4.7.2.4. 주의 사항

- 특정 작품명만 AFA\_PERFORMANCE로 태깅하고 '재즈댄스, 검무, 사교댄스, 에어로빅댄스' 등 춤 종류는 CV\_ART로 태깅한다.
  - 파소도블레는 라틴아메리카댄스 중 하나이다. → <파소도블레: AFA\_PERFORMANCE>는 <라틴아메리카댄스: CV\_ART> 중 하나이다.
- 뮤지컬은 연극 형식에 음악이 추가된 것이므로 AFA\_PERFORMANCE로 태깅하고, 오페라는 음악 형식에 춤이 추가된 것이므로 AFA\_MUSIC으로 태깅한다.
  - 오페라 '니벨룽의 반지'가 KBS1채널을 통해 여러분을 찾아갑니다. → 오페라 '<니벨룽의 반지: AFA\_MUSIC>'가 KBS1채널을 통해 여러분을 찾아갑니다.

### 4.7.3. AFA\_VIDEO

#### 4.7.3.1. 정의 및 설명

- 영상/라디오 작품명 및 TV/라디오 채널명
- 영상 작품의 하위 유형
  - TV 프로그램명
  - 라디오 프로그램명
  - 영화 작품명
  - 애니메이션 작품명
  - 기타 영상 작품명
- TV/라디오 채널의 하위 유형
  - TV 채널명
  - 라디오 채널명

#### 4.7.3.2. 분석 대상

- 영상 작품의 하위 유형
  - TV 프로그램명: 아침마당, EBS 클래식, 생활의 달인 등
  - 라디오 프로그램명: 두시탈출컬투쇼, 최화정의 파워타임, 배철수의 음악캠프 등
  - 영화 작품명: 기생충, 부산행, 죽은 시인의 사회 등
  - 애니메이션 작품명: 겨울왕국, 인어공주, 알라딘 등
  - 기타 영상 작품명: Feel the Rhythm of Korea: SEOUL, [공익광고협의회] 확산자 - 30초 등
- TV/라디오 채널의 하위 유형
  - TV 채널명: OCN, TBC, MBC 등
  - 라디오 채널명: MBC표준FM, SBS리브FM, TBS교통방송 등

#### 4.7.3.3. 예문

- TV 프로그램명
  - '<생방송! 팡팡 게임짱: AFA\_VIDEO>'은 개그맨 정용국과 게임자키 레이가 진행을 맡아 총 24회 분량이 방영될 예정이다.
- 라디오 프로그램명
  - 2월 17일 <배철수의 음악캠프: AFA\_VIDEO>는 30주년 기념 방송으로 진행된다.
- 영화 작품명
  - 세속적인 삶을 초월한 천재 화가를 '술에 취해 그림을 그리는 신선'으로 표현하여 영화의 제목은 <취화선: AFA\_VIDEO>이라고 붙였다.
- 애니메이션 작품명



- 원제는 《<바람의 검심-메이지 검객 낭만담: AFA\_VIDEO> (<るろうに剣心-明治剣客浪漫譚: AFA\_VIDEO>)》이다.
- 기타 영상 작품명
  - 한국관광공사가 제작한 이 전 세계에서 인기를 얻고 있다.
- TV 채널명
  - 4일 오후 7시 씨티극장에서 채널 <CGV: AFA\_VIDEO>를 통해 초청된 시청자를 포함한 관객들을 대상으로 ‘<친절한 금자씨: AFA\_VIDEO>’ 영화상영회를 갖고 박찬욱 감독과의 대화 시간도 마련한다.
- 라디오 채널명
  - ‘<두시탈출컬투쇼: AFA\_VIDEO>’는 <SBS파워FM: AFA\_VIDEO>의 간판 프로그램이다.

#### 4.7.3.4. 주의 사항

- 작품명의 일부가 절단되었지만 맥락을 통해 확인이 가능한 경우에도 해당 작품명으로 인정하여 AFA\_VIDEO로 태깅한다.
  - 요즘 마블 영화보다 핫한 김지영을 안 본 사람이 있을까요? → 요즘 마블 영화보다 핫한 <김지영: AFA\_VIDEO>을 안 본 사람이 있을까요?
- 뮤직비디오의 제목은 대부분 가수명과 곡명으로 이루어져 있으므로 각각을 PS\_NAME과 AFA\_MUSIC으로 태깅한다. 그러나 뮤직비디오가 하나의 영상 작품으로 언급된 경우에는 그 전체를 하나의 개체명으로 판단하여 AFA\_VIDEO로 태깅한다.
  - BTS의 'Dynamite' 새 뮤직비디오가 지난 주말에 공개되었다. → <BTS: PS\_NAME>의 <'Dynamite': AFA\_MUSIC> 새 뮤직비디오가 지난 주말에 공개되었다.
  - 'BTS(방탄소년단) Dynamite Official MV (B-side)'가 'BTS(방탄소년단) Dynamite Official MV'에 이어 유튜브 100만 뷰를 달성했다. → '<BTS(방탄소년단) Dynamite Official MV (B-side): AFA\_VIDEO>'가 '<BTS(방탄소년단) Dynamite Official MV: AFA\_VIDEO>'에 이어 유튜브 100만 뷰를 달성했다.

#### 4.7.4. AFA\_ART\_CRAFT

##### 4.7.4.1. 정의 및 설명

- 미술 작품명
- 미술 작품의 하위 유형
  - 그림의 제목
  - 조각의 제목
  - 기타 미술품의 제목

##### 4.7.4.2. 분석 대상

- 그림의 제목: 게르니카, 모나리자, 별이 빛나는 밤, 이제현 초상, 세한도 등
- 조각품의 제목: 지옥의 문, 생각하는 사람, 다비드, 경주 불국사 금동아미타여래좌상, 금동미륵보살반가사유상 등
- 기타 미술품의 제목: 금관총 금제 허리띠, 청자 오리모양 연적, 성덕대왕신종, 고려산적석사비, 광개토대왕릉비 등

#### 4.7.4.3. 예문

- 그림의 제목
  - 그 사례로서 단원 김홍도(檀園 金弘道)의 서명과 낙관이 찍힌 《<춘화사계첩: AFA\_ART\_CRAFT(春畵四季帖: AFA\_ART\_CRAFT)>》을 들 수 있다.
- 조각품의 제목
  - 《<지옥의 문: AFA\_ART\_CRAFT>》의 청동 작품은 모두 로댕의 사후에 제작된 것이다.
- 기타 미술품의 제목
  - 가까이 북한산에도 <북한산 진흥왕 순수비: AFA\_ART\_CRAFT>가 있다.

#### 4.7.4.4. 주의 사항

- 그림을 모은 화첩은 AFA\_DOCUMENT가 아니라 AFA\_ART\_CRAFT로 태깅한다.
  - 김홍도의 풍속화첩은 모두 여덟 면으로 구성되어 있다. → 김홍도의 <풍속화첩: AFA\_ART\_CRAFT>은 모두 여덟 면으로 구성되어 있다.
- “고려산적석사비(高麗山積石寺碑)”, “광개토대왕릉비” 등의 비(碑)와 그 비문도 미술품으로 다루어 AFA\_ART\_CRAFT로 태깅한다.
  - 드라마 태왕사신기는 광개토대왕릉비에 적힌 광개토대왕릉비문의 내용을 참고하여 만들었다고 한다. → 드라마 태왕사신기는 <광개토대왕릉비: AFA\_ART\_CRAFT>에 적힌 <광개토대왕릉비문: AFA\_ART\_CRAFT>의 내용을 참고하여 만들었다고 한다.

### 4.7.5. AFA\_MUSIC

#### 4.7.5.1. 정의 및 설명

- 음악 작품명
- 음악 작품의 하위 유형
  - 노래의 제목
  - 오페라 작품명
  - 뮤직비디오 작품명

#### 4.7.5.2. 분석 대상

- 노래의 제목: Dynamite, 좋은날, 미소 속에 비친 그대, 사의 찬미, 도술가, 서동요, 수궁가

등

- 오페라 작품명: 돈 조반니, 피가로의 결혼, 카르멘, 투란도트 등

#### 4.7.5.3. 예문

- 노래의 제목
  - BTS의 <Dynamite: AFA\_MUSIC>가 미국 빌보드 라디오 차트 톱 5에 진입했다.
  - 이날치의 <범 내려온다: AFA\_MUSIC>는 판소리 <수궁가: AFA\_MUSIC>의 한 대목으로, 육지에 올라온 자라가 호랑이를 처음 만나 놀라는 모습을 재미있게 노래하고 있다.
- 오페라 작품명
  - 《<운명의 힘: AFA\_MUSIC>》(1862), 《<돈 카를로스: AFA\_MUSIC>》(1867) 등과 더불어 1859년에 제작된 베르디의 중기 작품에 속한다.

#### 4.7.5.4. 주의 사항

- 유네스코 지정 한국의 무형유산인 '판소리'나 '종묘제례악'은 음악의 하위 장르로 볼 수 있으므로 CV\_ART로 태깅하고, 구체적인 작품명만 AFA\_MUSIC으로 태깅한다.
  - 국립창극단에서는 매년 판소리 수궁가가 공연된다. → 국립창극단에서는 매년 <판소리: CV\_ART> <수궁가: AFA\_MUSIC>가 공연된다.
- 노래의 제목 또는 가수명과 다르게 만들어진 앨범명은 AFW\_OTHER\_PRODUCTS로 태깅한다.
  - 푸른하늘의 3집 푸른하늘의 수록곡 푸른하늘 → <푸른하늘: PS\_NAME>의 <3집: QT\_ALBUM> <푸른하늘: AFW\_OTHER\_PRODUCTS>의 수록곡 <푸른하늘: AFA\_MUSIC>

### 4.8. AF\_WARES

- 판매되는 상품/제품의 명칭

#### 4.8.1. AFW\_SERVICE\_PRODUCTS

##### 4.8.1.1. 정의 및 설명

- 서비스업의 상품 명칭
- 서비스업의 상품 하위 유형
  - 금융 상품명
  - 여행 상품명
  - 교육 상품명

##### 4.8.1.2. 분석 대상

- 금융 상품명: 보험, 부동산, 예금, 적금, 펀드 등
  - 예금 상품명: 우리희망드림적금, 주택청약종합저축, 우리행복지킴이통장 등
- 여행 상품명: 패키지 여행, 호캉스, 에어비앤비 등
  - 패키지 여행 상품명: 동유럽 발칸 완전 일주 9일, 낭만과 열정의 멕시코 리비에라 크루즈 11일 등
- 교육 상품명: 강좌, 프로그램 등
  - 강좌명: 영화가 있는 저녁 무빙, 새콤달콤 과일 샌드위치3 등

#### 4.8.1.3. 예문

- 예금 상품명
  - 우리은행도 <하이-믹스(Hi-Mix) 복합예금 32호: AFW\_SERVICE\_PRODUCTS>를 500억 원 한도로 다음 달 3일까지 판매한다.
- 패키지 여행 상품명
  - '<동유럽 발칸 완전 일주 9일: AFW\_SERVICE\_PRODUCTS>' 상품은 매진이 얼마 남지 않았습니다.
- 교육 상품명
  - '<성경으로 배우는 라틴어: AFW\_SERVICE\_PRODUCTS>'는 지난주에 마감되었습니다.

#### 4.8.1.4. 주의 사항

- 금융상품 중 정확한 상품명에 쓰였을 경우에만 AFW\_SERVICE\_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 것들은 CV\_FUNDS로 태깅한다.
  - 우리은행은 노령연금을 받는 만 50세 이상 고객들을 대상으로 최대 연 1.4%의 우대이자율을 적용받을 수 있는 '시니어플러스 우리통장'을 판매한다. → 우리은행은 <노령연금: CV\_FUNDS>을 받는 만 50세 이상 고객들을 대상으로 최대 연 1.4%의 우대이자율을 적용받을 수 있는 '<시니어플러스 우리통장: AFW\_SERVICE\_PRODUCTS>'을 판매한다.

### 4.8.2. AFW\_OTHER\_PRODUCTS

#### 4.8.2.1. 정의 및 설명

- 식품과 음료를 제외한 공산품의 상품명
- 공산품의 하위 유형
  - 가구/인테리어용품명
  - 공업/농업용품명
  - 교육/사무용품명
  - 귀금속/보석명
  - 미용 및 위생용품명

- 스포츠/레저 용품명
- 신발/의류명
- 의료기기명
- 주방용품명
- 컴퓨터/가전제품명

#### 4.8.2.2. 분석 대상

- 가구/인테리어용품명: 책상, 침대, 의자, 액자, 꽃병, 각종 모형 등
  - 침대명: (시몬스) Beautyrest Zalmon, Beautyrest Edison, (일룸) 모션베드 캐주얼, (에이스) 아르노
- 공업/농업용품명: 페인트, 인조수지, 공구, 농기구 등
  - 농기구명: 콩콩이 식재삽, 해루질 3종 세트, 잡초킬러, 울산호미, 수원낫 등
- 교육/사무용품명: 인쇄물, 문방구 등
  - 사무용품명: (3M) 포스트잇 노트, (아모스) 딱풀, 방안직자 등
- 귀금속, 보석명: 보석 원석, 장신구, 기념품, 시계 등
  - 시계 제품명: (오메가) 씨마스터300, (롤렉스) 서브마리너 등
- 미용 및 위생용품명: 화장품, 향수, 치약, 샴푸, 염색약, 인조손톱, 청소용구 등
  - 화장품명: (아이오페) 모이스트젠 소프트너, (에스티로더) 어드밴스드 나이트 리페어, (겔랑) 메테오리트 펄 일루미네이팅 파우더 등
  - 향수 제품명: (샤넬) NO.5, (존 바바토스) 아티산, (다비도프) 쿨워터 맨, (불가리) 레젼메 코랄리아
- 스포츠/레저 용품명: 스포츠용구, 자전거, 등산, 낚시 캠핑용품, 완구 등
  - 자전거명: (코메트) 페인킬러, (삼천리) 메리다, (다혼) 비테세 등
  - 완구명: 뽀로로 흔들어리모컨, 콩순이 미니 펫샵, 파워레인저 엔진포스 전자검 등
- 신발/의류명: 운동화, 구두, 옷, 가방, 지갑, 장갑, 우산 등
  - 운동화 제품명: (나이키) 에어맥스97, 에어 조던 원, (뉴발란스) MX608BW1 등
  - 의류 제품명: (유니클로) 히트텍, (펜디) 피카부 등
- 의료기기명: 기저귀, 붕대, 마스크, 물티슈, 휠체어, 안마의자, 각종 병원용품 등
  - 물티슈명: (샤인) 그린터치 물티슈, (베베썬) 시그니처 그린 물티슈, (행복한) 행복한물티슈, (한울생약) 닥터스 손소독티슈 등
- 주방용품명: 접시, 컵, 조리도구, 각종 서빙용구 등
  - 커피잔명: (포트메리온) 테라스 커피잔, (웨지우드) 콜로세움 커피잔, (텐바이텐) 오슬로 에스프레소잔 등
- 컴퓨터/가전제품명: 게임, 프로그램, 냉장고, 에어컨 등 전자제품
  - 게임 상품명: 리니지M, AFK아레나, 리그 오브 레전드, 메이플스토리 등
  - 전자기기 제품명: 트롬(세탁기), 덩채(김치냉장고), 애니콜 가로본능, 아이폰SE2, 갤럭시Z

플립 등

#### 4.8.2.3. 예문

- 침대명
  - 일룸 관계자는 드라마 도깨비의 영향으로 '<모션베드 캐주얼: AFW\_OTHER\_PRODUCTS>'의 판매량이 급상승했다고 전했다.
- 농기구명
  - '<해루질 3종 세트: AFW\_OTHER\_PRODUCTS>'를 구매하시면 호미, 갈퀴 등을 따로 구매하는 번거로움을 피할 수 있습니다.
- 사무용품명
  - 아내는 매일 아침 출근 전, 오늘의 할 일이 적힌 '<포스트잇: AFW\_OTHER\_PRODUCTS>'을 냉장고에 붙여 놓곤 했다.
- 시계 제품명
  - 서른 살 생일을 맞아 롤렉스 '<서브마리너: AFW\_OTHER\_PRODUCTS>'를 일하느라 휴가도 못 간 저에게 선물했습니다.
- 화장품명
  - 에스티로더는 영원한 베스트셀러 '<어드밴스드 나이트 리페어: AFW\_OTHER\_PRODUCTS>'의 판매만으로도 화장품 업계 1위 자리를 올해도 지킬 수 있었다.
- 향수명
  - 샤넬 '<NO.5: AFW\_OTHER\_PRODUCTS>'는 '마릴린 먼로가 잠자리에 들 때 몸에 걸치는 유일한 옷'이라는 말로 유명하다.
- 완구명
  - '<콩순이 미니 펫샵: AFW\_OTHER\_PRODUCTS>', '<콩순이 말하는 냉장고: AFW\_OTHER\_PRODUCTS>', '<콩순이 자판기: AFW\_OTHER\_PRODUCTS>' 등 여자 아이들을 겨냥한 콩순이 시리즈 상품도 다양하다.
- 운동화 제품명
  - 90년대에 강남에서 중고등학교를 다닌 학생들에게 나이키 '<에어맥스97: AFW\_OTHER\_PRODUCTS>'은 또 하나의 교복이었다.
- 의류 제품명
  - 5월까지의 '<히트택: AFW\_OTHER\_PRODUCTS>'을 입어야 뼈마디에 바람이 안 들어.
- 물티슈명
  - 최근 '<행복한 물티슈: AFW\_OTHER\_PRODUCTS>'처럼 가성비 좋은 중소기업 상품에 관심을 기울이는 소비자들이 많다.
  - 커피잔명
    - 영국 왕실에서 사용한다고 하여 유명해진 웨지우드도 '<콜로세움: AFW\_OTHER\_PRODUCTS>', '<르네상스골드: AFW\_OTHER\_PRODUCTS>' 등의 제품으로 인기를 얻고 있다.

- 게임 상품명
  - '<리그 오브 레전드: AFW\_OTHER\_PRODUCTS>'가 10월 온라인게임 점유율 45%를 넘었다.
- 가전제품명
  - '<애니콜 가로본능: AFW\_OTHER\_PRODUCTS>'은 삼성전자에서 2004년에 출시한 휴대전화인 모델명 SCH-V500(SKT)의 다른 이름이다.

#### 4.8.2.4. 주의 사항

- 상품명에 브랜드명이 포함되어 있는 경우에도 상품명만 AFW\_OTHER\_PRODUCTS로 태깅하고 브랜드명은 OGG\_ECONOMY로 태깅한다.
  - 나이키에서 이번에 출시한 나이키 에어맥스97은 오리지널 디자인의 아름다움을 그대로 살리면서 기능성을 강화했다는 평가를 받는다. → <나이키: OGG\_ECONOMY>에서 이번에 출시한 <나이키: OGG\_ECONOMY> <에어맥스97: AFW\_OTHER\_PRODUCTS>은 오리지널 디자인의 아름다움을 그대로 살리면서 기능성을 강화했다는 평가를 받는다.
- 전자제품 중 정확한 상품명으로 나온 경우만 AFW\_OTHER\_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 제품들은 TM\_HARDWARE로 태깅한다.
  - 아이폰3GS는 한국의 스마트폰 시장을 흔들어 놓았다. → <아이폰3GS: AFW\_OTHER\_PRODUCTS>는 한국의 <스마트폰: TMI\_HW> 시장을 흔들어 놓았다.
- 의류, 스포츠용품 등 정확한 상품명에 쓰였을 경우에만 AFW\_OTHER\_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 것들은 CV로 태깅한다.
  - 등산복 중에서는 아이디어 '조셉 남성 아이스 쿨넥 폴로티'가 판매율 1위를 기록했다. → <등산복: CV\_CLOTHING> 중에서는 아이디어 '<조셉 남성 아이스 쿨넥 폴로티: AFW\_OTHER\_PRODUCTS>'가 판매율 1위를 기록했다.
- 별도의 항목으로 분류되어 있는 무기, 교통수단, 악기의 이름은 각각 AF\_MUSICAL\_INSTRUMENT, AF\_WEAPON, AF\_TRANSPORT로 태깅한다.
  - 지난 2011년 경매에 나온 '<레이디 블런트: AF\_MUSICAL\_INSTRUMENT>'는 기존 '<스트라디바리우스: AF\_MUSICAL\_INSTRUMENT>'의 경매가를 넘어서며 <바이올린: AF\_MUSICAL\_INSTRUMENT> 사상 최고가를 기록했다.
  - 그는 보직이 <기관총: AF\_WEAPON> 사수였기 때문에 '<K-2: AF\_WEAPON>'를 만져본 적도 없다고 했다.
  - 기아자동차는 '<카니발: AF\_TRANSPORT>' 4세대 출시를 앞두고 사전예약에 들어갔다.
- 'pansy pop, 밀라노쇼, 앤틱 웨지힐, 수줍은레드, 702호' 등 각 브랜드만의 색상 명칭이 제품명과 함께 사용되었을 경우에는 해당 제품명과 함께 묶어 AFW\_OTHER\_PRODUCTS로 태깅하며, 고유 색상 명칭만 사용되었을 경우에도 개체명의 일부로 해당 개체를 가리킨다고 판단하여 AFW\_OTHER\_PRODUCTS로 태깅한다.
  - 디올 립글로우 006 베리가 요즘 나의 최애템. → 디올 <립글로우 006 베리: AFW\_OTHE

R\_PRODUCTS>가 요즘 나의 최애편.

- 로지 코랄은 생얼 같은 자연스러움을 연출해 줍니다. → <로지 코랄: AFW\_OTHER\_PRODUCTS>은 생얼 같은 자연스러움을 연출해 줍니다.
  - 제품명은 FFLOW의 '오일수 새틴핏 미니 립스틱 04 로지코랄(Oilsoo Satin Fit Mini Lipstick 04 ROSY CORAL)'이지만 고유 색상 명칭인 '로지코랄'만으로도 특정 제품임을 드러낼 수 있으므로 AFW\_OTHER\_PRODUCTS로 태깅한다.

## 5. ORGANIZATION

---

ORGANIZATION은 기관 및 단체 명칭을 나타내며 경제 기관, 교육 기관, 군사 기관, 미디어 및 방송 기관, 스포츠 기관, 예술 기관, 의료 기관, 종교 기관, 과학 기관, 도서관, 법률 기관, 정치 및 정부 기관, 음식 관련 기관, 숙박 기관 등으로 세분화된다. 세부 기관에 포함되지 않는 기관 및 단체들은 OGG\_OTHERS로 태깅한다.

특정 기관 및 단체명과 그 기관 및 단체의 세부 조직 명칭이 동시에 쓰였을 경우, 이를 묶어 전체를 기관 및 단체명으로 보지 않고 대표 기관 및 단체명만을 태깅한다.

### 5.1. OGG\_ECONOMY

#### 5.1.1. 정의 및 설명

- 경제 관련 기관 및 단체
- 회사 및 기업
  - 공기업 및 사기업
  - 은행, 카드사, 보험사, 증권사 등
  - 개인 사업체 및 자영업
  - 영화사, 엔터테인먼트사
- 브랜드 이름

#### 5.1.2. 분석 대상

- 경제 관련 기관 및 단체
  - 금융감독원, 국제통화기금, 경제협력개발기구(OECD), 금융투자협회, 한국거래소, 한국신용평가
- 회사 및 기업
  - 공기업 및 사기업: 인천국제공항공사, 한국도로공사, 동양척식주식회사, 삼성전자, LG그



- 은행, 카드사, 보험사, 증권사: 카카오뱅크, 세계은행, 한국은행, 국민은행, 비씨카드, LG 화재, AIG, HMC 투자증권
- 개인 사업체 및 자영업: 멍스토랑, 뉴월드부동산, 열공 스테디카페, 제이엠 싱크대, 에펠 (수제화)
- 영화사, 엔터테인먼트사: CJ 엔터테인먼트, 굿프্রে임, 유니버설 스튜디오, JYP, SM 엔터테인먼트
- 브랜드 이름
  - 빈폴, 타미힐피거, 닥스, 에르메스, 구찌

- 경제 관련 기관 및 단체
  - <금융감독원: OGG\_ECONOMY>은 금융기관에 대한 검사를 요청할 수 있도록 하고 있다.
  - <국제통화기금: OGG\_ECONOMY>은 세계무역의 안정된 확대를 통하여 가맹국들의 고용 증대, 소득 증가, 생산자원개발에 기여하는 것을 궁극적인 목적으로 한다.
  - <금투협: OGG\_ECONOMY> 관계자는 "적립식 펀드는 대부분 주식형이어서 전체 주식형 펀드 판매 잔액 감소분의 70% 이상을 적립식 펀드가 차지한다"라며 "9월에도 적립식 펀드 판매 잔액 감소세는 계속 이어질 것으로 보인다"고 말했다.
  - <한국거래소: OGG\_ECONOMY>는 12일 오전 <한국거래소: OGG\_ECONOMY> 서울 사옥에서 옵티시스의 코스닥 신규 상장 기념식을 개최했다고 밝혔다.
  - <G20: OGG\_ECONOMY>이라는 것은 <G20: OGG\_ECONOMY>회원국들의 모임이기 때문에 IMF나 월드뱅크 같은 다른 국제기구들은 기술적인 지원을 해 주고 있습니다.
- 공기업 및 사기업
  - 현재 <한국도로공사: OGG\_ECONOMY>가 이를 위임 받아 도로의 유지관리를 전담하고 있는데 모두가 포장도로이다.
  - 1946년 <한흥메리야스: OGG\_ECONOMY>로 출발하여 1955년 <한흥산업(주): OGG\_ECONOMY>으로 상호를 바꾸었고 1957년 백양상표를 등록하였다. 1960년 <한흥물산(주): OGG\_ECONOMY>으로 상호를 변경하였으며, 1975년 기업을 공개하였다. 1977년 울마크를 획득하였고, 1979년 상호를 <(주) 백양: OGG\_ECONOMY>으로 변경하였다가, 1966년 다시 <(주)비와이씨: OGG\_ECONOMY>로 변경하였다.
  - 1일 오전 9시24분 현재 유가증권시장에서 <삼성전자: OGG\_ECONOMY>는 전 거래일 대비 175%(1만 3,000원) 오른 75만 7,000원에 거래 중이다.
  - <홀렛팩커드: OGG\_ECONOMY>(<Hewlett-Packard co.: OGG\_ECONOMY>)는 컴퓨터, 인터넷, 인트라넷솔루션 서비스, 통신제품 및 측정솔루션 등 여러 첨단정보사업 분야에서 탁월한 제품의 성능과 지원 면에서 세계적인 명성을 얻고 있는 초우량 글로벌기업이다.

- <동양척식주식회사: OGG\_ECONOMY>는 1908년에 일본이 조선의 토지와 자원을 수탈할 목적으로 설치한 식민지 착취기관이다.
- 은행, 카드사, 보험사, 증권사
  - 사실 <카카오뱅크: OGG\_ECONOMY> 계좌를 개설한 것은 요 체크카드가 갖고 싶어서 시작한 것인데 카드도 꽤 괜찮은 듯하다.
  - <세계은행: OGG\_ECONOMY>이 지난 22일 내년 세계경제 성장률을 2.8%로 예상한 만큼, 국제통화기금이 내년 전망치를 올릴 것이라는 정부의 예측도 일리가 있다.
- 개인 사업체
  - <부영이의 숲: OGG\_ECONOMY>이라는 부영이 카페 전단지를 받았다.
- 영화사, 엔터테인먼트사
  - <유니버설 픽처스: OGG\_ECONOMY>, 톰 크루즈와 더그 라이만 감독의 사상 '첫 우주로케이션 장편극 영화' 계약 논의 중.
  - <JYP엔터테인먼트: OGG\_ECONOMY>가 발표하는 곡마다 1위에 오르며 핫한 여름을 보내고 있다.
- 브랜드명
  - 공방패션인데..... <샤넬: OGG\_ECONOMY> 가방이랑.

#### 5.1.4. 주의 사항

- (태깅 대상) 기관 및 단체명은 공식적인 명칭에 한하며 '회사, 은행, 그룹, 보험사, 백화점' 등과 같은 보통명사를 제외한 구체적인 기관 및 단체명과 상호를 태깅 대상으로 한다.
  - 삼성전자에서 오늘 오후 신제품 발표가 있을 예정이다. → <삼성전자: OGG\_ECONOMY>에서 오늘 오후 신제품 발표가 있을 예정이다.
  - 오늘부터 모든 백화점들이 특별 할인 행사를 시작했다.
    - '백화점'은 보통명사로 태깅하지 않는다.
- (태깅 범위) 특정 기관 및 단체의 이름과 그 기관 및 단체의 세부 조직 명칭이 동시에 쓰였을 경우, 이를 묶어 전체를 기관 및 단체명으로 태깅하지 않고 대표 기관 및 단체명만을 태깅한다.
  - 공정거래위원회 사무처 → <공정거래위원회: OGG\_ECONOMY> 사무처
- (기관+지점) 신한은행 광화문점, 현대오일뱅크 대산공장, 교보문고 강남점 등과 같이 '기관+지점'이 함께 나올 경우 분리하여 태깅하지 않고 전체를 묶어서 하나로 태깅한다.
  - 신한은행 광화문점 → <신한은행 광화문점: OGG\_ECONOMY>
  - 현대오일뱅크 대산공장 → <현대오일뱅크 대산공장: OGG\_ECONOMY>
- (경제 관련 공공 기관 및 정부 기관) 경제 관련 기관 중 공공 기관 및 정부 기관은 OGG\_POLITICS로 태깅한다.
  - 근로복지공단, 한국산업단지공단 → <근로복지공단: OGG\_POLITICS>, <한국산업단지공단: OGG\_POLITICS>

- (산업단지) '구로공단'과 같은 산업단지는 OGG\_ECONOMY로 태깅하지 않고 LC\_OTHERS로 태깅한다.
  - 구로공단 → <구로공단: LC\_OTHERS>
- (G20) 'G20'과 같은 경제 모임은 OGG\_ECONOMY로 태깅하나 'G20 정상회의, 다보스포럼' 등과 같은 경제 회의는 EV\_OTHERS로 태깅한다.
  - <G20: OGG\_ECONOMY>이라는 것은 <G20: OGG\_ECONOMY>회원국들의 모임이기 때문에 IMF나 월드뱅크 같은 다른 국제기구들은 기술적인 지원을 해 주고 있습니다.
  - 오늘 이 자리에 <G20 정상회의: EV\_OTHERS>에서 다양한 자원봉사를 했던 분들이 함께 하고 있습니다.
- (방위산업체) 특정한 기업을 지칭하는 것이 아니기 때문에 개체명 대상이 아니므로 태깅하지 않는다.
- (중복 개체명) 기관 및 단체가 OG, LC, AF로 중복 태깅이 가능할 경우 문맥과 술어를 고려하여 태깅한다. '~에 다니다, ~을 경영하다, ~이 흥하다, ~이 잘 되다, ~을 차리다'와 같은 술어와 함께 쓰이면 OG로 태깅하고 '~에서 만나다, ~에서 구입하다'와 같은 술어와 함께 쓰일 경우 LC\_OTHERS로 태깅한다. 또한 '~이 보인다/흔들리다, ~을 짓다/만들다'와 함께 쓰이며 건축물의 의미로 해석되는 경우 AF\_BUILDING으로 태깅한다. 그러나 술어를 고려하여 태깅하되 술어의 다의성을 주의하여 태깅한다.
  - 이 사장이 <롯데백화점: OGG\_ECONOMY>을 경영하면서 매출이 점점 증가하고 있다.
  - 우리는 주말에 <롯데백화점: LC\_OTHERS>에서 만나기로 했다.
  - 명동에 있는 <롯데백화점: AF\_BUILDING>은 1979년에 지어졌다.
- (개체명의 일부로 해당 개체를 가리키는 경우) 회사, 은행, 상호명 등이 텍스트 내 다른 문장, 문맥을 고려하지 않고도 단독으로 줄임말로 쓰이는 경우 (보편적으로 줄임말로 자주 사용되는 경우) 복원된 형태에 준하여 태깅한다.
  - <국민: OGG\_ECONOMY>, <신한: OGG\_ECONOMY> 등 5대 시중 은행의 8월 말 중소기업 대출 잔액은 223조 4,364억 원으로 전달보다 1조 4,935억 원 감소했다.

## 5.2. OGG\_EDUCATION

### 5.2.1. 정의 및 설명

- 교육 기관 및 단체
  - 학교
  - 학원
  - 훈련소
- 평생 교육 기관

- 학회

### 5.2.2. 분석 대상

- 교육 기관
  - 학교: 밀알 유치원, 서천초등학교 병설 유치원, 잠원중학교, 경희고등학교, 광주인화학교, 포스텍, 경희사이버대학교, 서울대학교, 차의과학대학, 국군간호사관학교, 중앙직업전문학교, 국자감
  - 학원: 청담어학원, 라이프치히 음악원, 엘리트 화랑태권도, 종로학원
  - 훈련소: 고용훈련소, 논산훈련소, 소방학교, 광주특수전교육단
  - 평생 교육 기관: 메가존아이티평생교육원
- 교육 단체
  - 한국과학교육단체, 경제교육단체협의회
- 학회
  - 국제한국어교육학회, 한국환경교육학회, 조선지리학회

### 5.2.3. 예문

- 교육 기관
  - 1949년 <홍익대학: OGG\_EDUCATION>으로 개편, 1961년 대학정비령에 따라 일시 <홍익미술대학: OGG\_EDUCATION>으로 칭하였다가 이듬해 <홍익대학: OGG\_EDUCATION>으로 환원하였으며
  - 경상남도 진주에서 태어나 1950년 <경남중학: OGG\_EDUCATION>을 졸업하고, 1954년 <부산대학교: OGG\_EDUCATION> 법학과를 졸업하였으며, 1964년에는 <서울대학교: OGG\_EDUCATION> 행정대학원에서 석사학위를 취득했다.
  - 1980년 1월 <국군간호사관학교: OGG\_EDUCATION>로 명칭을 변경하고 간호전문대학 과정으로 승격하였다.
- 교육 단체
  - <경제교육단체협의회: OGG\_EDUCATION>가(25개 기관) 시행하고 있는 교육 내용을 홍보하고, 다양한 프로그램을 체험할 수 있도록 구성
- 학회
  - <국제한국어교육학회: OGG\_EDUCATION>: 한국어 교육, 제20권, 제30호

### 5.2.4. 주의 사항

- (태깅 대상) 학교, 학원, 초등학교 등과 같은 보통명사는 태깅하지 않는다.
- (학과, 단과대학) OGG\_EDUCATION으로 태깅하지 않는다.
  - 1954년 부산대학교 법학과를 졸업하였으며 → 1954년 <부산대학교: OGG\_EDUCATION> 법학과를 졸업하였으며

- '부산대학교 법학과'는 '법학과'를 빼도 '부산대학교'라는 큰 의미는 변화가 없다.
- '법학과'는 태깅 대상이 아니지만 '법학, 국어국문학' 등은 FD로 태깅한다.
- 홍익대 미대 → <홍익대: OGG\_EDUCATION> 미대
  - 학과와 단과대는 학교에 포함이 되어 있는 개념이기 때문에 태깅하지 않는다.
  - 다만 '~도서관'의 경우 앞에 사람 이름이 나오고 도서관이 나온 경우 사람 이름만 태깅을 하면 안 된다. 사람 이름과 도서관은 개체명 태그상 완전히 다른 의미이며, 도서관을 뺐을 때 '~도서관'이라는 개체명의 의미가 사라진다. 반면 '부산대학교 법학과'는 '법학과'를 빼도 '부산대학교'라는 큰 의미는 변화가 없다. 앞의 단어와 뒤의 단어가 포함관계인지를 고려하는 것이 필요하다.
- (대학교+OO캠퍼스) '한양대학교 서울캠퍼스'와 같이 '대학+OO캠퍼스'의 경우 분리하지 않고 하나로 묶어서 태깅한다.
  - 한양대학교 서울캠퍼스 → <한양대학교 서울캠퍼스: OGG\_EDUCATION>
  - 경희대학교 국제캠퍼스 → <경희대학교 국제캠퍼스: OGG\_EDUCATION>
- (훈련소) 훈련소는 모두 OGG\_EDUCATION으로 태깅한다. '고용훈련소'와 같이 특정 이름이 부여되지 않은 경우라도 OGG\_EDUCATION으로 태깅한다.
  - 고용훈련소 → <고용훈련소: OGG\_EDUCATION>
  - 훈련소 → <훈련소: OGG\_EDUCATION>
- (유테스코, 유엔) 교육, 문화, 사회, 정치, 과학 등의 여러 분야와 관련된 기관은 OGG\_OTHERS로 태깅한다.
- (학회) 학회의 종류에 OGG\_EDUCATION 이외의 세부적인 태그가 가능한 경우에는 OGG 세부 분야로 태깅하고, 세부 태그가 없는 경우에는 OGG\_EDUCATION으로 태깅한다.
  - 국제한국어교육학회 → <국제한국어교육학회: OGG\_EDUCATION>
  - 한국정보과학회 → <한국정보과학회: OGG\_SCIENCE>
- (군사 교육 기관 및 훈련소) '웨스트포인트사관학교' 등의 군사학교와 '광주특수전교육단' 등의 군사 교육을 시키는 기관도 OGG\_EDUCATION으로 태깅한다.

## 5.3. OGG\_MILITARY

### 5.3.1. 정의 및 설명

- 군사 기관 및 단체
- 군사 유형
- 테러 조직 및 무장 단체

### 5.3.2. 분석 대상

- 군사 기관 및 단체

- 국방품질관리소, 한미연합사령부, 국방홍보원, 국제연합군, 대한민국재향군인회, 주한미군, 31사단, 일본군
- 군사 유형
  - 해군, 공군, 육군, 인민군
- 테러 조직 및 무장 단체
  - 탈레반, 알카에다

### 5.3.3. 예문

- 군사 기관 및 단체
  - 5월 20일 유엔 사무총장의 조정교섭이 실패로 돌아가자 <영국군: OGG\_MILITARY>은 포클랜드에 상륙, 75일 간의 격전 끝에 6월 14일 <아르헨티나군: OGG\_MILITARY>의 항복으로 전쟁을 종결시켰다.
  - 19일 광주 동부경찰서가 조선대 교수 이메일 해킹에 현역 군인이 연루돼, 사건을 <육군: OGG\_MILITARY> <31사단: OGG\_MILITARY>으로 넘겼다면서 <31사단: OGG\_MILITARY> 헌병대와 경찰이 합동으로 수사에 착수했다고 말했다.
  - 양국의 합의로 설치된 <한미군사위원회: OGG\_MILITARY>의 전략지시에 의해 <한미연합사령부: OGG\_MILITARY>가 창설되었다.
  - 2000년 8월 18일 <국방홍보원: OGG\_MILITARY>으로 명칭이 바뀌었다.
- 군사 유형
  - 어제 오후 강원도 홍천에서는 기동훈련을 위해 이동하던 <육군: OGG\_MILITARY> 모 부대 소속 장갑차가 앞서가던 장갑차를 들이받았습니다.
  - <해군: OGG\_MILITARY>은 사고 원인 규명을 위해 고속정 승조원에 대한 조사와 함께 침몰 고속정 인양을 검토하고 있습니다.
- 테러 조직 및 무장 단체
  - <알카에다: OGG\_MILITARY>의 테러공격이 이달 말에 있을 것이라는 구체적인 정보가 입수되면서 독일에 테러비상이 내려졌습니다.

### 5.3.4. 주의 사항

- (독립운동 단체) 독립운동 단체 중 전투 단체는 OGG\_MILITARY로 태깅한다.
  - 오적암살단, 대한독립군 → <오적암살단: OGG\_MILITARY>, <대한독립군: OGG\_MILITARY>
- (군함) 군함은 문맥에 따라 AF\_TRANSPORT와 OGG\_MILITARY로 태깅한다. 문맥적으로 판단이 불가능할 경우 AF\_TRANSPORT로 태깅한다.
  - 한국의 구축함으로는 <광개토대왕함: AF\_TRANSPORT>과 <양만춘함: AF\_TRANSPORT> 등이 있다.
  - <광개토대왕함: OGG\_MILITARY>은 경기도 구리시와 자매결연을 맺고 정기적으로 지역 봉사활동에 참여하기로 했다.

- (군사 학교, 군사 교육 및 훈련 기관) 해군사관학교와 같은 군사 학교와 광주특수전교육단과 같은 군사 교육 기관은 OGG\_EDUCATION으로 태깅한다.
  - 해군사관학교, 광주특수전교육단 → <해군사관학교: OGG\_EDUCATION>, <광주특수전교육단: OGG\_EDUCATION>

## 5.4. OGG\_MEDIA

### 5.4.1. 정의 및 설명

- 미디어 기관 및 단체
- 방송 관련 회사 및 기업

### 5.4.2. 분석 대상

- 미디어 기관 및 단체
  - KBS, MBC, TJB, 불교 방송, 기독교 방송, 한국일보, 씨네21, 국제잡지협회, 방송심의위원회, 공익광고협의회, 월스트리트저널, 로이터통신, AP통신, 조선중앙통신, 코엑스 미디어 센터
- 방송 관련 회사 및 기업
  - 스카이라이프(주)

### 5.4.3. 예문

- 미디어 기관 및 단체
  - 국립국어원과 <SBS: OGG\_MEDIA>는 10월12일 오후 4시 서울 <SBS: OGG\_MEDIA> 목동사옥에서 제6회 방송 언어 공동 연구 발표회를 개최한다.
  - <로이터통신: OGG\_MEDIA>은 26일 마타이 박사가 케냐 수도 나이로비 병원에서 전날 향년 71세로 사망했다고 케냐 언론을 인용해 보도했다.
  - 24일 문형진 통일교 세계회장과 세계평화연합 조의 방문단이 평양에 도착했다고 <조선중앙통신: OGG\_MEDIA>이 보도했다.
  - 1980년까지는 <방송심의위원회: OGG\_MEDIA>의 업무를 <한국방송윤리위원회: OGG\_MEDIA>가 수행하였다.
- 방송 관련 회사 및 기업
  - <스카이라이프(주): OGG\_MEDIA>는 2002년 12월 19일 위성방송 사업권을 획득한 이후 2001년 3월부터 본격적인 위성방송 준비를 시작하여

### 5.4.4. 주의 사항

- (신문사) '조선일보' 등과 같은 신문사가 기관의 의미를 나타낼 때는 OGG\_MEDIA로 태깅하

고 출판물의 의미를 나타낼 때는 AFA\_DOCUMENT로 태깅한다.

- 사무실에 무단 침입해 서류를 촬영하다 서울시로부터 고발당한 조선일보 출입기자에 대한 징계수위를 정하기 위해 → 사무실에 무단 침입해 서류를 촬영하다 서울시로부터 고발당한 <조선일보: OGG\_MEDIA> 출입기자에 대한 징계수위를 정하기 위해
- 오늘 아침 조선일보는 참 읽을거리가 많았다. → 오늘 아침 <조선일보: AFA\_DOCUMENT>는 참 읽을거리가 많았다.
- (전문지, 잡지) 출판물이 아닌 기관의 의미로 쓰일 경우 OGG\_MEDIA로 태깅한다.
  - 애니메이션 전문지: 애니메이션툰
  - 영화 잡지: 씨네21

## 5.5. OGG\_SPORTS

### 5.5.1. 정의 및 설명

- 스포츠 기관 및 단체
- 스포츠 팀 및 구단 이름
- 스포츠 리그
- 인터넷게임 팀

### 5.5.2. 분석 대상

- 스포츠 기관 및 단체
  - 국민체육진흥심의위원회, 국기원, 국민생활체육협의회, 한국빙상지도자협의회, 국군체육부대
- 스포츠 팀 및 구단 이름
  - 한국팀, 중국팀, 축구대표팀, 아이스하키 대표팀, 경찰청 야구단, 시카고 화이트삭스, LA 다저스, 한화 야구단, 맨체스터유나이티드
- 스포츠 리그
  - K 리그, 프리미어 리그, 메이저 리그
- 인터넷게임 팀
  - 아프리카 프릭스, 삼성전자 Khan

### 5.5.3. 예문

- 스포츠 기관 및 단체
  - 1991년 1월 사단법인 <국민생활체육협의회: OGG\_SPORTS>가 창립총회를 거쳐 정식 출범하였다.
  - 1984년 <체육지도대: OGG\_SPORTS>를 모태로 <국군체육부대: OGG\_SPORTS>를 재창



설하고 전군(全軍)의 체육종목을 통합하였다.

- 1996년 8월10일 <비상회보> 창간호를 발간하고, 2002년 10월 1일 <한국빙상지도자협의회: OGG\_SPORTS> 홈페이지를 개설하였다.
- 스포츠 팀 및 구단 이름
  - 오늘 같은 경우도 그렇고 이경훈 선수가 3일, 네 번째 경기에서 점수를 충분히 줄이면서 <한국팀: OGG\_SPORTS>에 힘을 실어주고 있습니다.
  - 우리 <축구대표팀: OGG\_SPORTS>이 중국을 3:0으로 가볍게 물리치고 8강에 진출했습니다.
- 스포츠 리그
  - <메이저리그: OGG\_SPORTS> 공식 홈페이지가 추신수는 한국을 아시안게임 챔피언으로 이끌며 병역 혜택을 받았다고 비중 있게 다루었습니다.
- 인터넷게임 팀
  - <아프리카 프리क्स: OGG\_SPORTS>의 미드 라이너 '플라이' 송용준은 서머 시즌 준수한 활약을 펼쳤다.

#### 5.5.4. 주의 사항

- (태깅 대상) '~대표팀'과 같은 보통명사도 OGG\_SPORTS로 태깅한다.
  - 양궁 대표팀 → <양궁 대표팀: OGG\_SPORTS>
- (국가 대표팀) 앞에 특정한 스포츠 유형이 들어가야만 태깅할 수 있다.
  - 한국 축구 국가 대표팀 → <한국: LC\_COUNTRY> <축구 국가 대표팀: OGG\_MEDIA>
  - 한국 국가 대표팀 → <한국: LC\_COUNTRY> 국가 대표팀
- (리그) 1부(리그) 등은 '잉글랜드 3부 리그'로 구체화된 경우에만 OGG\_SPORTS로 태깅한다.
- (우승팀, 준우승팀) 우승팀, 준우승팀과 같은 보통명사는 태깅하지 않는다.
- (운동 시설) 골프장, 수영장 등과 같은 운동 시설이 건물이 아닌 기관의 의미로 쓰일 때에는 OGG\_SPORTS로 태깅한다.
  - <한강 공원 수영장: OGG\_SPORTS>은 많은 피서객들을 유치하기 위해 시설 정비에 들어갔다.
  - <에이 스크린 골프 연습장: OGG\_SPORTS>은 최신 카카오프렌즈 스크린을 도입했다.

### 5.6. OGG\_ART

#### 5.6.1. 정의 및 설명

- 예술 기관 및 단체
- 공연장, 전시장, 박물관, 미술관, 극장(기관의 의미로 쓰인 경우로 한정)

### 5.6.2. 분석 대상

- 예술 기관 및 단체
  - 독일디자인협회, 마린스키발레단, 로스앤젤레스 필하모니 관현악단
- 공연장, 전시장, 박물관, 미술관, 극장
  - 국립중앙박물관, 루브르박물관, 세종문화회관, 예술의 전당, 리움미술관, 대한극장

### 5.6.3. 예문

- 예술 기관 및 단체
  - 기아차는 20일 <독일디자인협회: OGG\_ART>에서 올해 처음으로 자동차 부문 우수 브랜드 선정을 위해 실시한 '2011 오토모티브 브랜드 콘테스트'에서 브랜드 디자인 부문 최우수상을 수상했다고 밝혔다.
  - 1919년 윌리엄 A.클라크 2세(William A. Clark, Jr.)가 창단한 <로스앤젤레스 필하모니 관현악단: OGG\_ART>은 예술제와 강연에도 참여하고 있으며, 모든 연령층의 청중에게 무료 음악회를 열어 주기도 한다.
- 공연장, 전시장, 박물관, 미술관, 극장
  - <예술의 전당: OGG\_ART>은 LG 유플러스와 문화예술 콘텐츠 보급 독점 업무 협약을 맺었다.
  - 프랑스 <루브르박물관: OGG\_ART>이 전시 그림 구입을 위해서 <루브르박물관: OGG\_ART> 역사상 최초로 일반인을 대상으로 모금운동을 벌이고 있습니다.

### 5.6.4. 주의 사항

- (태깅 대상) 공연장, 전시장, 박물관, 영화관 등의 보통명사는 태깅하지 않는다.
- (공연장, 전시장, 박물관 등) 기관이 아닌 건축물의 의미로 사용되었다고 판단되는 경우, OG가 아닌 AF로 태깅한다. '~를 운영하다, ~를 경영하다, ~이 망하다'와 같은 서술어와 어울리면 OGG\_ART로 태깅하고 '~이 열린다'와 같은 서술어와 어울리면 AF\_BUILDING으로 태깅한다.
  - LG아트센터는 '명성황후', '영웅'의 윤호진 연출과 손잡고 신작 '보이책'을 제작해 10월 9일부터 한 달간 무대에 올린다. → <LG아트센터: OGG\_ART>는 '명성황후', '영웅'의 윤호진 연출과 손잡고 신작 '보이책'을 제작해 10월 9일부터 한 달간 무대에 올린다.
  - 30일 서울 LG아트센터에서는 카운터 테너 필립 자루스키와 베니스 바로크 오케스트라 공연이 열립니다. → 30일 서울 <LG아트센터: AF\_BUILDING>에서는 카운터 테너 필립 자루스키와 베니스 바로크 오케스트라 공연이 열립니다.

## 5.7. OGG\_MEDICINE

### 5.7.1. 정의 및 설명

- 의학·의료 기관 및 단체
- 병원, 약국

### 5.7.2. 분석 대상

- 의학·의료 기관 및 단체
  - 세계보건기구, KMI한국의학연구소, 한국원자력의학원, 대한간호협회, 의료보험연합회
- 병원, 약국
  - 삼성서울병원, 국립의료원, 국립재활원, 아주대병원, 센정형외과, 경희한의원, 새봄약국

### 5.7.3. 예문

- 의학·의료 기관 및 단체
  - 1948년 <대한간호협회: OGG\_MEDICINE>로 개칭하였고, 1994년에는 <국제간호협회의: OGG\_MEDICINE>(ICN)에 정회원국으로 가입하였다.
  - 1988년 <의료보험연합회: OGG\_MEDICINE>로 명칭을 변경하여 공단의 진료비 심사기구를 흡수, 통합하였으며
  - 중국의 유력한 차기 지도자 시진핑 부주석의 부인이자, "국민 가수"로까지 불리는 미모의 여인 핑리위안이 최근 <세계보건기구: OGG\_MEDICINE>의 후천성면역결핍증(AIDS) 예방 친선대사에 임명됐다.
- 병원, 약국
  - <삼성서울병원: OGG\_MEDICINE> 부인암 센터에서는 오는 10월 18일 '부인암 환자의 날' 행사를 개최한다.
  - 1986년 8월 18일 <국립재활원: OGG\_MEDICINE>으로 직제를 개정하였으며, 1994년 1월 29일 원내에 재활병원을 설치하였다.

### 5.7.4. 주의 사항

- (태깅 대상) 병원, 약국, 동물 병원 등과 같은 보통명사는 태깅하지 않는다. 구체적인 대상의 이름이 나온 경우에 한하여 태깅한다.
  - 이대목동병원 인공관절센터 김영호 소장 → <이대목동병원: OGG\_MEDICINE> 인공관절센터 김영호 소장
  - 집 근처에는 동물병원이 없어서 → 집 근처에는 동물병원이 없어서
    - 구체적인 병원의 이름이 없기 때문에 태깅하지 않는다.
- (의학 관련 학문 분야 및 분과) 정형외과학, 한의학, 이비인후과, 정형외과 등은 FD\_MEDICINE으로 태깅하고 '센정형외과'와 같이 병원을 나타낼 때는 OGG\_MEDICINE으로 태깅한다.
  - 그렇다면 정형외과와 재활의학과를 체험하고 → 그렇다면 <정형외과: FD\_MEDICINE>와

- <재활의학과: FD\_MEDICINE>를 체험하고
- 췌장형외과에서 다리 수술을 했다. → <췌장형외과: OGG\_MEDICINE>에서 다리 수술을 했다.

## 5.8. OGG\_RELIGION

### 5.8.1. 정의 및 설명

- 종교 기관 및 단체
- 종교 및 종파

### 5.8.2. 분석 대상

- 종교 기관 및 단체
  - 사랑의 교회, 명동 성당, 대한예수교장로회, 기독교한국침례회, 대한성공회, 대한불교조계종, CCC
- 종교 및 종파
  - 증산교, 구가톨릭교, 기독교, 불교, 천주교, 이슬람, 힌두교, 원불교, 장로교, 성결교, 화엄종

### 5.8.3. 예문

- 종교 기관 및 단체
  - <사랑의 교회: OGG\_RELIGION>가 오는 14일까지 18차 특별 새벽 부흥 예배 '세계 교회와 함께 하는 말씀과 부흥의 현장'을 진행합니다.
  - <대한예수교장로회: OGG\_RELIGION> 통합 총회가 회복을 위한 첫걸음으로 대각성 기도회를 개최했습니다.
  - <대한불교조계종: OGG\_RELIGION>이 올해 초 핵심 화두로 내걸었던 자성과 쇄신운동을 본격적으로 추진하기 위해 지난달 30일 "자성과 쇄신 결사 추진본부"를 설립하고 본부장에 도법 스님을 임명했다.
  - 뮌헨에서 구가톨릭교도 대회를 열고 독립교단을 발족시켰으며 오스트리아, 스위스, 프랑스, 이탈리아 등 여러 나라에도 <구가톨릭 교회: OGG\_RELIGION>가 생겨났다.
- 종교 및 종파
  - 가입 상담을 요청하자 직원이 가지고 온 것은 <불교: OGG\_RELIGION> 상조회 상품 안내 책자였습니다.
  - <기독교: OGG\_RELIGION> 단체가 나경원 한나라당 서울시장 후보를 지지하는 이메일을 무작위로 발송해 서울시 선거관리위원회가 불법선거 운동 혐의로 조사에 착수했다.

#### 5.8.4. 주의 사항

- (복합명사) '구가톨릭교도'와 같이 종교 기관이 복합명사의 일부이고, 종교 기관만을 독립적인 의미로 분리하기 힘든 경우에는 OGG\_RELIGION으로 태깅하지 않는다.
- (유교) 종교 단체가 아니라 철학사상, 일종의 학문이므로 FD\_HUMANITIES로 태깅한다.
- YMCA는 청소년의 인격과 지도력 훈련 교육 및 청소년 문화사업을 중점으로 하는 기관으로 OGG\_OTHERS로 태깅한다.

### 5.9. OGG\_SCIENCE

#### 5.9.1. 정의 및 설명

- 과학 기관 및 단체
- 연구소 및 연구원

#### 5.9.2. 분석 대상

- 과학 기관 및 단체
  - 국제전기통신연합, 한국전기안전공사, 정보통신연구진흥원, 한국발명진흥회, 한국인터넷진흥원
- 연구소 및 연구원
  - 한국표준과학연구원, 한국에너지기술연구원, 전기안전시험연구원, 국립수산물과학원, 독도수산물연구센터

#### 5.9.3. 예문

- 과학 기관 및 단체
  - 1999년 법정법인인 <정보통신연구진흥원: OGG\_SCIENCE>으로 새로 출범한 뒤 그해 해외 정보통신기술협력센터와 부설 정보통신기술이전센터를 신설하였다.
  - 교육과학기술부는 성공적인 나로호 3차 발사를 위해 각계 전문가 15명으로 나로호 <발사점검위원회: OGG\_SCIENCE>를 구성하고 본격 활동에 착수한다.
- 연구소 및 연구원
  - <한국표준과학원: OGG\_SCIENCE> 분석화학표준센터 김인중 박사는 체내 필수 영양소인 셀레늄을 정확하게 측정할 수 있는 기술을 개발했다고 27일 밝혔다.
  - <한국에너지기술연구원: OGG\_SCIENCE>(원장 황주호)은 기존에 난방이나 급탕용으로만 이용하던 태양열을 활용해 미래의 에너지원이라 불리는 수소를 생산하는 기술을 개발했다고 17일 밝혔다.
  - <독도연구센터: OGG\_SCIENCE>가 '해양과학으로 보는 동해와 독도'란 주제로 심포지엄을 개최한다.

#### 5.9.4. 주의 사항

- (과학 관련 학회) '한국정보과학회'와 같은 과학 관련 학회도 OGG\_SCIENCE로 태깅한다.

### 5.10. OGG\_LIBRARY

#### 5.10.1. 정의 및 설명

- 도서관
- 도서관 관련 기관 및 단체

#### 5.10.2. 분석 대상

- 도서관
  - 국립중앙도서관, 국회도서관, 바티칸도서관, 법학전문도서관, 일송도서관, 광천공공도서관
- 도서관 관련 기관 및 단체
  - 한국도서관협회

#### 5.10.3. 예문

- 도서관
  - 외규장각도서를 한국에 돌려주는 데 대해 <국립도서관: OGG\_LIBRARY> 등 일부 프랑스 문화계가 강하게 반발하고 나섰습니다.
  - <수성구립용학도서관: OGG\_LIBRARY>은 21일 오후 7시 시청각실에서 서정시 낭송 작은 음악회를 개최한다.
  - <양구도서관: OGG\_LIBRARY>이 오는 14, 15일 이틀간 평생 학습 축제를 개최한다.
  - 1471년부터는 <바티칸도서관: OGG\_SCIENCE>을 확장하는 등 학문과 예술을 장려함으로써 문화사상(文化史上) 위대한 업적을 남겼다.
  - <용산도서관: OGG\_LIBRARY> 맞은편 <남산도서관: OGG\_LIBRARY> 옥상공원은 입구 철문이 잠겨 있다.
- 도서관 관련 기관 및 단체
  - 1955년 <한국도서관협회: OGG\_LIBRARY>가 재발족하면서 도서관 교육 문제가 거론되고, 대학에서도 도서관학을 교과과정에 넣게 되었으며, 1957년부터 도서관학이 대학교의 독립된 과(科)로 인정받게 되었다.

#### 5.10.4. 주의 사항

- (태깅 대상) '도서관'과 같은 보통명사는 태깅하지 않는다.
- (사람 이름+도서관) 사람 이름과 도서관을 함께 태깅한다.

- 김대중도서관 → <김대중도서관: OGG\_LIBRARY>

## 5.11. OGG\_LAW

### 5.11.1. 정의 및 설명

- 법률 기관 및 단체

### 5.11.2. 분석 대상

- 법률 기관 및 단체
  - 서울중앙지법, 국제형사재판소, 광주지법, 서울행정법원, 대법원
  - 대한법률구조공단, 대한법무사협회, 저작권심의조정위원회, 대한변호사협회

### 5.11.3. 예문

- 법률 기관 및 단체
  - <특검팀: OGG\_LAW>은 오늘(7일) 오전 이재환 전 삼성벤처투자 사장을 불러 이재용 전 무가 주도한 인터넷 사업 실패를 삼성 계열사에 떠넘겼다는 의혹을 받고 있는 '이 삼성' 사건에 대해 조사하고 있다.
  - 경북대병원 순환기내과 조모(48) 교수는 최근 공무원연금공단 이사장을 상대로 공무상요양불승인처분 취소 청구 소송을 <서울행정법원: OGG\_LAW>에 제기했다.
  - <헌법재판소: OGG\_LAW>는 법적 분쟁을 사법적으로 해결한다는 점에서는 일반법원과 같으나 특히 정치적 파급효과가 큰 헌법적 분쟁을 대상으로 하며
  - 저작권에 관한 사항을 심의하고 분쟁을 조정하기 위해 <저작권 심의조정위원회: OGG\_LAW>를 둔다.
  - 변호사로서 개업을 하고자 하는 때에는 지방변호사회를 거쳐 <대한변호사협회: OGG\_LAW>에 등록을 하여야 한다.

### 5.11.4. 주의 사항

- (법원) 법원은 국가 기관(OGG\_POLITICS)과 법률 기관(OGG\_LAW)에 모두 해당되는데 OGG\_LAW로 태깅한다.
- (국가명+도시명+법원) 국가명은 따로 태깅하고 도시명과 법원은 함께 태깅한다.
  - 대한제국 법부 전라북도재판소 → <대한제국: LCP\_COUNTRY> <법부: OGG\_LAW> <전라북도재판소: OGG\_LAW>
- (국가 기관+법률 기관) 국가 기관과 법률 기관을 따로 태깅한다.
  - 미 군정청 광주지방법원 → <미 군정청: OGG\_POLITICS> <광주지방법원: OGG\_LAW>

## 5.12. OGG\_POLITICS

### 5.12.1. 정의 및 설명

- 정부 및 행정 기관
- 공공 기관
- 정치 기관

### 5.12.2. 분석 대상

- 정부 및 행정 기관
  - 청와대, 국회, 대사관, 미래부, 정보통신부, 동사무소, 시청, 군청, 경제기획원, 대검, 검찰청, 정부, 우체국, 경찰서, 소방서
- 공공 기관
  - 근로복지공단, 사립학교교직원연금공단, 한국사회적직업진흥원
- 정치 기관
  - 새누리당, 더불어민주당, 독립협회

### 5.12.3. 예문

- 정부 및 행정 기관
  - <검찰: OGG\_POLITICS>과 <경찰: OGG\_POLITICS>이 <청와대: OGG\_POLITICS>와 <국방부: OGG\_POLITICS> 등 주요 국가기관의 인터넷 홈페이지를 마비시킨 해킹사건에 대한 공조수사에 착수했다.
  - 그러나 1999년 5월 정부조직법이 다시 개정되어 <기획예산처: OGG\_POLITICS>가 신설되면서 <예산청: OGG\_POLITICS>은 <기획예산처: OGG\_POLITICS>의 예산실로 축소 개편되었다.
  - 개정으로 <내무부: OGG\_POLITICS>에서 <행정자치부: OGG\_POLITICS>로 명칭이 변경되었다.
  - 10대 여중생이 장난 삼아 어린 아이들의 발을 걸어 넘어뜨리게 한 CCTV 영상이 공개돼 <경찰: OGG\_POLITICS>이 진상파악에 나섰습니다.
  - 중국 <정부: POLITICS>는 지난 월요일 G20 서울정상회의의 참가와 관련해 특별 기자회견을 열었습니다.
- 정치 기관
  - <한나라당: OGG\_POLITICS>은 <민주당: OGG\_POLITICS>의 예산심사 거부가 직무유기라며 예결위 정책질의에 착수했습니다.

### 5.12.4. 주의 사항

- (독립운동 기관) 일종의 정치 기관으로 고려해서 OGG\_POLITICS로 태깅한다.



- 간도 국민회, 간도대한민회 → <간도 국민회: OGG\_POLITICS>, <간도대한민회: OGG\_POLITICS>
  - 단, 독립운동 단체 중에 전투 단체는 OGG\_MILITARY로 태깅한다.
- (국가 연구기관) OGG\_POLITICS로 태깅한다.
  - 집현전
    - 국가연구기관이지만 과학 기관은 OGG\_SCIENCE로 태깅한다.
- (경찰) 행정기관으로 OGG\_POLITICS를 의미하기도 하고, 경찰 공무원으로 CV\_OCCUPATION을 의미하기도 하는데 그 판단 기준은 서술어를 중심으로 문장 의미에 따라 결정한다. '~에 민원을 제기하다'와 같은 술어와 같이 쓰일 경우에 OGG\_POLITICS로 태깅하고 '~으로 일하다, ~해서 생계를 유지하다'와 같은 술어와 같이 쓰일 경우에는 CV\_OCCUPATION으로 태깅한다.
  - '교통'이라고 적힌 연두색 조끼를 입은 경찰을 배치했다. → '교통'이라고 적힌 연두색 조끼를 입은 <경찰: CV\_OCCUPATION>을 배치했다.
  - 경찰에 신고하다. → <경찰: OGG\_POLITICS>에 신고하다.
- (국가명 및 도시명) 맥락에 따라 LC가 될 수도 있고 OG가 될 수도 있는데, 그 판단 기준은 해당 문장의 서술어를 중심으로 문장 의미에 따라 결정한다. '~사람, ~인'과 같은 명사와 같이 쓰일 경우에 LC로 태깅하고 '~으로 귀화하다, ~으로 망명하다'와 같은 술어와 같이 쓰일 경우 또한 LC로 태깅한다.
  - <러시아: LCP\_COUNTRY>, <프랑스: LCP\_COUNTRY> 같은 전쟁한 잠수함 강국들은 물론 '스승'이라 할 <독일: LCP\_COUNTRY>까지 물리치고 거둔 성과여서 더욱 대견스럽다.
  - <프랑스: OGG\_POLITICS>도 '방송장악' 논란
  - 한편 <고양시: OGG\_POLITICS>는 내년 10월에 열리는 제92회 전국체전도 유치했다.
  - <고양시: LCP\_CITY> 덕양구 행신동 서정초등학교 인근 아파트형 공장과 관련해
- (조선시대 정치 집단) 서론, 노론 등 조선시대 정치 집단도 OGG\_POLITICS로 태깅한다.
- (국가+정부), (국가+도시+정부 기관), (사람 이름+정부)와 같은 복합 명사구는 각각 따로 태깅한다.
  - 한국 정부 → <한국: LCP\_COUNTRY> <정부: OGG\_POLITICS>
  - 대한민국 상해 임시정부 → <대한민국: LCP\_COUNTRY> <상해: LCP\_CITY> <임시정부: OGG\_POLITICS>
  - 이명박 정부 → <이명박: PS\_NAME> <정부: OGG\_POLITICS>
- (LC+OG) LC가 단순히 기관이 소재한 장소를 의미할 경우에는 각각 태깅한다. 그러나 LC를 포함하여 하나의 기관 이름으로 기능할 경우에는 전체를 함께 OG로 태깅한다.
  - 미국 컬럼비아대학 → <미국: LCP\_COUNTRY> <컬럼비아대학: OGG\_EDUCATION>
    - 컬럼비아 대학이 미국에 위치한다는 뜻이므로 나누어 태깅한다.
  - 남구를 지구당 위원장 → <남구를: LC\_OTHERS> <지구당: OGG\_POLITICS> <위원장: CV\_POSITION>

- 광주교육청 → <광주교육청: OGG\_POLITICS>
  - 교육청이 광주에 위치한다는 의미가 아니라 광주 지역의 교육 행정을 책임지는 기관이라는 의미이며, 전체가 하나의 기관으로 인식되기 때문에 통합하여 태깅한다.
  - '한국 정부'와 '광주교육청'의 태깅 방식이 다른 것은 띄어쓰기와 관련이 없다. '광주교육청'의 경우 개체의 실체가 명확하며 문맥에 따라 AF로도 볼 수 있으나 '한국 정부'는 개체의 실체가 명확하지 않고 'OO 정부'처럼 여러 정부의 종류를 지칭하는 것으로 많이 쓰이기 때문에 <한국: LCP\_COUNTRY> <정부: OGG\_POLITICS>로 태깅한다.
- (줄임 표현) '시, 군, 구'가 줄임 표현으로 나올 시 기관의 의미를 나타내면 OGG\_POLITICS로 태깅한다.
  - <양구군: OGG\_POLITICS>은 (중략) <군: OGG\_POLITICS>에서
    - '시, 군, 구'가 단독으로 나와도 기관의 의미를 가질 경우에는 OGG\_POLITICS로 태깅한다.
- (임시회, 정기회) 기관과 연관된 단어가 아니며, 특정한 의미를 가지지 않는 포괄적인 단어이기 때문에 개체명 태깅 대상이 아니다.
- (합의제행정기관, 하급행정기관) 특정한 의미를 가지는 기관이 아니므로 개체명 태깅 대상이 아니다.

## 5.13. OGG\_FOOD

### 5.13.1. 정의 및 설명

- 음식 관련 업체 및 회사

### 5.13.2. 분석 대상

- 음식 관련 업체 및 회사
  - 본죽, 맥도날드, 김밥천국, 맛나분식, 던킨도너츠, 스타벅스, 파리바게트, 파스쿠찌, 도미노 피자

### 5.13.3. 예문

- 음식 관련 업체 및 회사
  - 아이스크림 전문점 <배스킨라빈스: OGG\_FOOD>는 오는 9월 이달의 새 맛으로 초콜릿과 피넛버터를 결합한 '땅콩 로켓' 아이스크림을 선보인다고 30일 밝혔다.
  - <뚜레쥬르: OGG\_FOOD>가 지난달 15일 빵 28종을 평균 8% 올린 데 이어 케이크류 가격을 최고 30% 인상했다.
  - <스타벅스: OGG\_FOOD>가 막대형 커버 '비아(Via)' 출시를 통해 국내 인스턴트 커피 시

장에 본격 진출한다.

- 우선 저희 집은 경기도 포천에서 작은 <바위식당: OGG\_FOOD>이라는 식당을 운영하고 있습니다.

#### 5.13.4. 주의 사항

- (태깅 대상) 본죽, 맥도날드, 스타벅스 등이 OGG\_ECONOMY로 여겨지기도 하지만 모든 음식과 관련된 업체 및 회사는 모두 OGG\_FOOD로 태깅한다.
- (상호명+OO점) '스타벅스 홍대점'과 같이 '상호명 + OO점'이 나올 경우 하나로 묶어서 태깅한다.
  - 스타벅스 홍대점 → <스타벅스 홍대점: OGG\_FOOD>
- (OO점) 개체명의 일부로 해당 개체를 가리키는 경우, 앞의 문맥에서 구체적인 상호명을 알 수 있는 경우에 한하여 태깅한다.
  - 와 보고 싶었던 한림점. 비오는 엔트러사이트도 나름 괜찮네.
    - '한림점'이 지칭하는 구체적인 대상을 앞의 문맥에서 알 수 있는 경우 태깅을 하고 구체적인 대상을 알 수 없는 경우에는 태깅하지 않는다.

### 5.14. OGG\_HOTEL

#### 5.14.1. 정의 및 설명

- 숙박 관련 업체 및 회사
- 호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장

#### 5.14.2. 분석 대상

- 숙박 관련 업체 및 회사
- 호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장
  - 신라호텔, 한화콘도, 대명리조트, 위커힐, 호텔신라, 그랜드인터컨티넨탈호텔, 백담사산장

#### 5.14.3. 예문

- 호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장
  - 외교통상부는 10월 19일 세종연구소 및 조선일보와 공동으로 <신라호텔: OGG\_HOTEL> (영빈관)에서 한일중 3국 국제학술회의를 개최한다.
  - 한국기업평가에 이어 한국신용평가도 <호텔롯데: OGG\_HOTEL>의 신용등급을 한 단계 상향 조정했다.
  - 흥사관광단지를 지나 <유선여관: OGG\_HOTEL>과 대흥사 파안교를 건너면 오른쪽 언덕에서 석간수가 솟아나고 왼쪽으로 대흥사 석주가 우뚝 솟아 보인다.

- 단, 이번 주말에는 <휘닉스파크: OGG\_HOTEL>에서만 가능하고 다음 주에는 수도권 지역에서만 가능합니다.

#### 5.14.4. 주의 사항

- (중복 태깅 우선순위) 휴양시설은 OGG\_HOTEL과 LC\_OTHERS가 모두 가능한데 OGG\_HOTEL을 우선하여 태깅한다.
- (산장) 산장의 경우 구체적인 상호명과 함께 쓰였을 경우 태깅한다. '백담사산장'과 같이 여관으로 등록되어 있고 상호명으로 쓰일 경우 전체를 OGG\_HOTEL로 태깅한다. 단, 백담사에 있는 산장의 의미로 쓰일 경우에는 '백담사'만 태깅하고 '산장'은 태깅하지 않는다.
- (룸 이름) 호텔 내 존재하는 룸 이름은 태깅하지 않는다.
  - 롯데호텔 주니어 스위트룸 → <롯데호텔: OGG\_HOTEL> 주니어 스위트룸

### 5.15. OGG\_OTHERS

#### 5.15.1. 정의 및 설명

- 기관 및 단체의 세부 분류에 해당되지 않는 나머지 모든 기관 및 단체

#### 5.15.2. 분석 대상

- 유색인종향상전국협회, 국제소비자연맹
- 육영재단, 녹색연합, 참여연대, 한국정신대문제대책협의회, 청목회
- 부모님 전상서팀, 울산의 남매팀
- 팬클럽 이름: 블링크, 원잇(ONE IT), 아미
- 게임 관련 기관
  - 길드, 클랜, 파티

#### 5.15.3. 예문

- <육영재단: OGG\_OTHERS>이란 넓은 의미에서는 후세 교육을 위해 설립한 재단에 대한 총칭으로 쓰이기도 한다.
- 결성 당시에는 <참여민주사회와 인권을 위한 시민연대: OGG\_OTHERS>였으나 두 차례 변경을 거쳐 <참여연대: OGG\_OTHERS>로 확정되었다.
- 1990년 11월 16일 일본군 중군위안부(정신대)에 대한 진실규명과 어렵게 사는 일본군 중군위안부 생존자들을 지원하려는 37개 여성·시민·종교·학생단체들이 모여 만든 연합단체이다. 줄여서 <정대협: OGG\_OTHERS>이라 한다.
- <특별한 인연팀: OGG\_OTHERS>부터 <울산의 남매팀: OGG\_OTHERS>까지 다섯 팀이 정말 다양한 사연과 다양한 노래를 불러주셨습니다.

- 김 지사는 오늘 <관훈클럽: OGG\_OTHERS> 초청 토론회에서 개헌은 지금 시점에 가능하지도 않고 바람직하지 않다면서도 청와대와 국회의 확실한 분권과 지방 자치 실현을 위해서는 제도를 법률적으로 바꾸면 된다고 말했습니다.

#### 5.15.4. 주의 사항

- (세부 분류가 애매한 기관 및 단체) '비세그라 그룹' 등의 세부 분류가 애매한 협력체는 OGG\_OTHERS로 태깅한다.
- (UN, 유네스코) 교육, 문화, 정치, 사회, 과학 등의 여러 분야와 관련된 기관으로 OGG\_OTHERS로 태깅한다.

## 6. LOCATION

LOCATION은 지역/장소와 지형/지리 명칭 등을 모두 포함한다. 지역/장소는 LCP(LoCation Place) 계열로 표시하고, 지형/지리 명칭은 LCG(LoCation Geography) 계열로 표시한다. 그 외 천체 명칭은 LC\_SPACE, LC 계열의 세부 유형이 아닌 기타 장소는 LC\_OTHERS로 태깅한다. 속담이나 게임, 가상의 지역/장소, 지형/지리 명칭은 태깅하지 않는다.

- 지역/장소명은 주로 공식적 지역 명칭과 행정구역 명칭을 나타낸다.
- 지역/장소명은 각각 분리하여 태깅한다.
  - 충청남도 예산군 덕산면 상가리 37 → <충청남도: LCP\_PROVINCE> <예산군: LCP\_COUNTY> <덕산면: LCP\_COUNTY> <상가리: LCP\_COUNTY> <37: QT\_ADDRESS>
- 지리 명칭은 국가명이나 행정구역명과는 다른 산, 강과 같은 자연 지리 명칭을 나타낸다.
- 지형 명칭에 지형의 유형은 포함되지 않고, 구체적인 지형 명칭만 해당된다.
  - <오세아니아: LCG\_CONTINENT>는 세계에서 가장 작은 대륙입니다.

### 6.1. LCP\_COUNTRY

#### 6.1.1. 정의 및 설명

- 국가명 및 왕국의 명칭

#### 6.1.2. 분석 대상

- 국가명

- 대한민국, 일본, 중국, 미국, 영국, 터키, 그리스, 짐바브웨, 케냐 등
- 왕국
  - 부탄 왕국, 오만 왕국, 통가 왕국 등

### 6.1.3. 예문

- 국가명
  - 동부의 리제는 <터키: LCP\_COUNTRY> 전 지역 중에서 가장 강수량이 많아 연평균 1,788mm에 달한다.
  - 예술은 <그리스: LCP\_COUNTRY>의 생활을 잘 반영하고 있다. 여름이면 고대 극장에서 <그리스: LCP\_COUNTRY>의 연극을 상연한다.
  - <케냐: LCP\_COUNTRY>는 여러 인종이 혼재하는 다인종(多人種) 사회의 특색을 지니고 있다.
- 왕국
  - 불교 국가인 <부탄: LCP\_COUNTRY>은 군인보다 승려가 많다고 한다.
  - <통가: LCP\_COUNTRY>에 대한 첫인상은 '이상한 나라의 앨리스' 속 비밀의 정원 같은 곳이었다.

### 6.1.4. 주의 사항

- (중복 개체명) 문맥상 '국가'를 의미할 때만 LCP\_COUNTRY로 태깅한다.
  - 이 대통령은 이날 오전 10시 반부터 청와대에서 방한 중인 알리벤 봉고온딤바 가봉 대통령과 정상회담 일정 등을 소화했다. → 이 대통령은 이날 오전 10시 반부터 청와대에서 방한 중인 알리벤 봉고온딤바 <가봉: LCP\_COUNTRY> 대통령과 정상회담 일정 등을 소화했다.
  - 韓美의 주요 관계자들은 한국의 평화를 위해 노력하기로 합의했다. → <韓: OGG\_POLITICS><美: OGG\_POLITICS>의 주요 관계자들은 <한국: LCP\_COUNTRY>의 평화를 위해 노력하기로 합의했다.
  - 우즈베키스탄이 카타르에 비기거나 지면 한국은 조 2위로 본선에 오르게 된다. → <우즈베키스탄: OGG\_SPORTS>이 <카타르: OGG\_SPORTS>에 비기거나 지면 <한국: OGG\_SPORTS>은 조 2위로 본선에 오르게 된다.
  - 문맥상 정부를 의미할 때는 OGG\_POLITICS, 스포츠 팀을 의미할 때는 OGG\_SPORTS로 태깅한다.
  - 조선에서 바로 일본 식민 지배로 넘어갔고 해방되자마자 소련 공산당의 영향권에 들어갔다. → <조선: DT\_DYNASTY>에서 바로 <일본: LCP\_COUNTRY> 식민 지배로 넘어갔고 해방되자마자 <소련: LCP\_COUNTRY> 공산당의 영향권에 들어갔다.
  - '왕조/시대'를 의미할 때는 LCP\_COUNTRY로 태깅하지 않는다.
- 국가를 약칭(略稱)으로 표현한 경우도 태깅한다.

- 韓美 국민들의 행복지수 비교 → <韓: LCP\_COUNTRY><美: LCP:COUNTRY> 국민들의 행복지수 비교
- 민주공화국은 특정한 나라가 앞에 붙어 있는 경우만 개체명 대상으로 인식한다.
  - 콩고는 지금의 민주공화국이 되기 위해 험난한 일들을 겪었다. → <콩고: LCP\_COUNTRY>는 지금의 민주공화국이 되기 위해 험난한 일들을 겪었다.
  - 자이르는 콩고민주공화국의 전 이름이다. → <자이르: LCP\_COUNTRY>는 <콩고민주공화국: LCP\_COUNTRY>의 전 이름이다.
    - 구체적인 국가명이 아닌 '국가/ 나라/민주공화국' 등의 보통명사는 태깅하지 않는다.

## 6.2. LCP\_PROVINCE

### 6.2.1. 정의 및 설명

- 도, 주 지역명

### 6.2.2. 분석 대상

- 도
  - 경기도, 강원도, 충청남도 등
- 주
  - 캘리포니아주, 애리조나주, 퀘벡주, 슈비츠주, 네바다주 등

### 6.2.3. 예문

- 도
  - 전국의 행정 구역을 13도로 구획하면서 <충청남도: LCP\_PROVINCE>와 <충청북도: LCP\_PROVINCE>로 나누어, 대체로 현재와 같은 행정 구역의 윤곽을 지니게 되었다.
  - 이번 태풍으로 <경남: LCP\_PROVINCE>과 <경북: LCP\_PROVINCE>은 큰 피해를 입었다.
- 주
  - 이 <퀘벡주: LCP\_PROVINCE>가 캐나다 연방에서 이탈하여 독립하려는 운동이 1960년대 부터 일기 시작하여 1980년에 처음으로 분리를 두고 주민의 찬반투표가 이루어졌으나 반대 60%로 무산되었다.
  - 스위스 중부에 있는 <슈비츠주: LCP\_PROVINCE>의 주도로 역사적인 건축물이 많다.
  - 최대산업은 관광산업으로, <네바다주: LCP\_PROVINCE>는 거의 모든 관광 · 오락에 대해 장려하며, 각지에 인공적인 관광 · 오락시설이 발달되었다.

### 6.2.4. 주의 사항

- 나라에 따라서 행정 구역의 세분화된 단위가 다르므로, 도시보다 큰 단위의 모든 행정 구

역은 LCP\_PROVINCE로 태깅한다.

- 빅토리아 시티를 홍콩의 수도로 알고 있는 사람이 있는데, 정확하게는 홍콩의 행정적 중심지 역할을 하는 지역이다. → 빅토리아 시티를 <홍콩: LCP\_PROVINCE>의 수도로 알고 있는 사람이 있는데, 정확하게는 <홍콩: LCP\_PROVINCE>의 행정적 중심지 역할을 하는 지역이다.
  - '홍콩, 할리우드'와 같이 특별자치구역이나 대도시권 등은 '도시보다 큰 단위의 행정 구역'이므로 LCP\_PROVINCE로 우선 태깅한다. '홍콩시, 할리우드시'와 같이 문맥상 '시'가 확실할 경우만 LCP\_CITY로 태깅한다.
- 중국이나 일본의 지역명도 태깅 범위에 포함된다. 'OO성'도 우리나라의 '도'에 해당하는 행정 구역이기 때문에 LCP\_PROVINCE로 태깅한다.
  - 길림성은 중국 북동부에 위치하고 있으며 지린성이라고도 한다. → <길림성: LCP\_PROVINCE>은 중국 북동부에 위치하고 있으며 <지린성: LCP\_PROVINCE>이라고도 한다.
- 뒤에 '지방/지역'이 붙으면 '지방/지역'을 제외하고 태깅한다.
  - 호남지방에 내린 폭우 → <호남: LCP\_PROVINCE> 지방에 내린 폭우
    - '호남/영남'과 같이 특정 지역명이 아닌 '중부/남부 지역'과 같이 지역의 부분을 나타내는 경우 '중부/남부'만 TM\_DIRECTION으로 태깅한다. '중부권/남부권'과 같이 'OO권'으로 사용된 경우 복합어이기 때문에 TM\_DIRECTION으로 내부 태깅하지 않는다.

## 6.3. LCP\_COUNTY

### 6.3.1. 정의 및 설명

- 군, 구, 동, 면, 읍, 리 등과 같은 세부 행정 구역명

### 6.3.2. 분석 대상

- 군, 구, 동
  - 영월군, 완도군, 종로구, 강서구, 노원구, 노암동, 방화동, 서천동 등
- 면, 읍, 리
  - 도암면, 춘포면, 담양읍, 양양읍, 횡계리, 영남리 등

### 6.3.3. 예문

- 군, 구, 동
  - 1986년 영암·강진·해남·장흥의 4개 군에 속했던 도서를 통합하여 <완도군: LCP\_COUNTY>이라 칭하였다.
  - <종로구: LCP\_COUNTY>의 관할 구역은 <사직동: LCP\_COUNTY>, <삼청동:



LCP\_COUNTY>, <부암동: LCP\_COUNTY>, <평창동: LCP\_COUNTY> 등 17개 동으로 정부 기관이 밀집되어 있다.

- 면, 읍, 리
  - <담양읍: LCP\_COUNTY> 일원에서는 가내수공업으로 죽세공예품을 생산한다.
  - 서쪽으로 <공덕면: LCP\_COUNTY>(<孔德面: LCP\_COUNTY>), 남쪽으로 <용지면: LCP\_COUNTY>(<龍池面: LCP\_COUNTY>), 북쪽으로 익산시 <춘포면: LCP\_COUNTY>(<春浦面: LCP\_COUNTY>)에 접한다.
  - 1952년 12월 북한의 행정 구역 개편 때 면이 폐지되면서 통합되어 신설한 <은천군: LCP\_COUNTY> <영남리: LCP\_COUNTY>가 되었다가 1974년 5월 남포직할시에 이관되었다.

#### 6.3.4. 주의 사항

- 나라에 따라서 행정 구역의 세분화된 단위가 다르므로, 도시보다 작은 단위의 모든 행정 구역은 LCP\_COUNTY로 태깅한다.
  - 이나바 감독은 12일 일본 도쿄도(東京都) 미나토구(港区)에 있는 그랜드프린스호텔에서 → 이나바 감독은 12일 일본 도쿄도(東京都) <미나토구: LCP\_COUNTY>(<港区: LCP\_COUNTY>)에 있는 그랜드프린스호텔에서
  - 지난 주말엔 친구들과 신촌에서 모였어요! → 지난 주말엔 친구들과 <신촌: LCP\_COUNTY>에서 모였어요!
    - '신촌, 강남'과 같이 도시 안의 일부 지역을 나타낼 때는 '도시보다 작은 단위의 행정 구역'이므로 LCP\_COUNTY로 우선 태깅한다. '신촌역, 강남역'과 같이 문맥상 '역'이 확실할 경우만 LC\_OTHERS로 태깅한다.
- 홍콩, 할리우드와 같이 도시보다 큰 구역은 LCP\_PROVINCE로 태깅하고, 신촌과 같이 도시보다 작은 구역은 LCP\_COUNTY로 태깅한다.
- 중국이나 일본의 지역명도 태깅 범위에 포함된다. 'OO현'은 우리나라의 '도'보다 작은 개념의 행정 구역이기 때문에 LCP\_COUNTY로 태깅한다.
  - 일본의 나가사키현은 유명한 음식이 많다. → 일본의 <나가사키현: LCP\_COUNTY>은 유명한 음식이 많다.
- 2014년부터 정식 시행된 새로운 주소 표기법인 도로명주소법에는 도로명이 주소로 표시된다. 도로명주소법에서의 도로명은 AF\_ROAD로 태깅한다.
  - 서울특별시 강서구 금남화로 154 → <서울특별시: LCP\_CAPITALCITY> <강서구: LCP\_COUNTY> <금남화로: AF\_ROAD> <154: QT\_ADDRESS>
- 행정 구역상 정식 명칭일 경우만 LCP\_COUNTY로 태깅한다. 선거구를 나타내는 'OO구를 /OO구갑' 등의 경우 행정 구역상 정식 명칭이 아니므로 LC\_OTHERS로 태깅한다.
  - 서울시 노원구는 노원갑, 노원을, 노원병 등 모두 세 명의 지역구 의원이 있다. → 서울특별시 <노원구: LCP\_COUNTY>는 <노원갑 : LC\_OTHERS>, <노원을: LC\_OTHERS>, <노

- 원병: LC\_OTHERS> 등 모두 세 명의 지역구 의원이 있다.
- (중복 개체명) 지역 이름과 역 이름이 동일할 경우 문맥에 따라 LCP\_COUNTY나 LC\_OTHERS로 태깅한다. 문맥으로 판단이 어려운 경우 LCP\_COUNTY로 우선 태깅한다.
    - 휴가 기간 동안 시부야 인근에서 계속 있었다. → 휴가 기간 동안 <시부야: LCP\_COUNTY> 인근에서 계속 있었다.
    - 혜화 근처에 3중 추돌 사고가 나서 좀 늦었어. → <혜화: LCP\_COUNTY> 근처에 3중 추돌 사고가 나서 좀 늦었어.
  - '나주'와 같이 과거에는 '군'이었으나 현재 '시'인 도시는 LCP\_CITY로 태깅한다. 문맥상 확실히 'OO군'일 경우만 LCP\_COUNTY로 태깅한다.
    - 고려·조선시대를 통하여 주로 목(牧), 부(府)로서 전남지역 행정의 중심지를 이루었으며, 1895년 <나주군: LCP\_COUNTY>이 되었다.
    - 과거나 현재나 <나주: LCP\_CITY>는 여전히 전남 최고의 농산물 집산지이다.

## 6.4. LCP\_CITY

### 6.4.1. 정의 및 설명

- 도시명

### 6.4.2. 분석 대상

- 도시
  - 강릉, 부산, 태백, 뉴욕 등
  - 맨체스터, 덴버, 요코하마, 부에노스아이레스, 이스탄불 등

### 6.4.3. 예문

- 도시
  - 현재 <맨체스터: LCP\_CITY>는 공업도시보다는 상업·금융·보험 등의 상업도시로서 런던에 버금가는 도시가 되었다.
  - 제임스 W.덴버(James W. Denver)의 이름을 따 <덴버: LCP\_CITY>로 지명을 바꾸었다.
  - 혼모쿠곶과 오오카강 하구 사이의 천해(淺海)가 매축되어 현재의 <요코하마: LCP\_CITY>의 중심부가 형성되고,
  - 터키 최대의 공업도시이기도 한 <이스탄불: LCP\_CITY>은 섬유·식품가공·농기구·고무 등의 공업이 발달하였으며 군수공업도 중요한 몫을 차지한다.

### 6.4.4. 주의 사항

- '동대구/서대구'와 같이 도시의 일부를 나타내는 '동/서/남/북+도시명'인 경우 도시보다 작

은 구역이므로 LCP\_COUNTY로 태깅한다. 뒤에 '지방/지역'이 붙으면 '지방/지역'을 제외하고 태깅한다.

- 이번 장마로 동대구 지역이 물에 잠겼다. → 이번 장마로 <동대구: LCP\_COUNTY> 지역이 물에 잠겼다.
  - 문맥상 '동대구역'을 의미할 경우 LC\_OTHERS로 태깅한다. 문맥으로 판단이 어려운 경우 LCP\_COUNTY로 우선 태깅한다.
- 집값 상승이 두드러져 전통적인 주거 선호지역인 동부산권이 압승을 거뒀습니다. → 집값 상승이 두드러져 전통적인 주거 선호지역인 동부산권이 압승을 거뒀습니다.
  - '서부산권, 동대구권'과 같이 '방향+도시+권'일 경우 복합어이기 때문에 내부 태깅하지 않는다.
- 국가의 수도명은 LCP\_CAPITALCITY으로 태깅한다.
  - 부에노스아이레스는 1536년에 창건되었으나 인디오의 공격을 받아 1541년에 방치 되었다가, 1580년에 재건되었다. → <부에노스아이레스: LCP\_CAPITALCITY>는 1536년에 창건되었으나 인디오의 공격을 받아 1541년에 방치 되었다가, 1580년에 재건되었다.

## 6.5. LCP\_CAPITALCITY

### 6.5.1. 정의 및 설명

- 수도명

### 6.5.2. 분석 대상

- 수도
  - 서울, 북경, 베이징, 도쿄, 런던 등
  - 아테네, 워싱턴, 캔버라, 파리, 브라질리아 등

### 6.5.3. 예문

- 수도
  - <아테네: LCP\_CAPITALCITY>는 고대 그리스 아티카 지방의 중심지였다.
  - 1790년 국회의 동의를 얻어 미국의 수도로 지정되었고, 1878년 조지타운이 <워싱턴: LCP\_CAPITALCITY>의 일부가 된 이래 새로운 국가의 영구 행정부 소재지로 선정된 컬럼비아 특별구와 공존하고 있다.
  - 수도 <캔버라: LCP\_CAPITALCITY>와 주변지역으로 이루어져 있다.
  - '수도예의 집중현상'은 프랑스의 특징이며, <파리: LCP\_CAPITALCITY>는 세계 제4위의 인구밀집 지역이다.
  - <브라질리아: LCP\_CAPITALCITY>를 21세기의 도시로 건설하였다.

#### 6.5.4. 주의 사항

- 예전에 수도였더라도 현재 없어졌거나 일반 도시면 LCP\_CITY로 태깅한다. 현재 수도인 도시만 LCP\_CAPITALCITY로 태깅한다.
  - 아르마비르, 음반자콩구, 경주 → <아르마비르: LCP\_CITY>, <음반자콩구: LCP\_CITY>, <경주: LCP\_CITY>
  - 당시 신라의 수도인 <경주: LCP\_CAPITALCITY>에서 대규모 전투가 벌어졌다.
    - 현재 수도가 아니라도 문맥상 확실히 수도를 나타내는 경우는 LCP\_CAPITALCITY로 태깅한다. 문맥으로 판단이 어려운 경우 LCP\_CITY로 우선 태깅한다.

### 6.6. LCG\_RIVER

#### 6.6.1. 정의 및 설명

- 강, 호수/육수의 이름

#### 6.6.2. 분석 대상

- 강
  - 섬진강, 대동강, 한강 등
- 호수/육수
  - 충주호, 동강, 춘천호, 아산호 등
  - 청계천, 중랑천, 탄천 등
  - 관곡지, 예당저수지 등

#### 6.6.3. 예문

- 강
  - <동강: LCG\_RIVER>은 완택산(完澤山:916 m)과 곰봉(1,015 m) 사이의 산간지대를 감입 곡류하며 남서쪽으로 흐르다가 영월군 영월읍 하송리(下松里)에서 <서강: LCG\_RIVER> (<西江: LCG\_RIVER>)과 만나 <남한강: LCG\_RIVER> 상류로 흘러든다.
- 호수/육수
  - 기암괴봉이 거대한 병풍처럼 펼쳐지면서 <충주호: LCG\_RIVER>와 어우러져 뛰어난 경관을 연출한다.
  - <춘천호: LCG\_RIVER>에 담긴 물은 <북한강: LCG\_RIVER> 상류 지역을 호수화하여 멀리 화천군 <파로호: LCG\_RIVER>(<破虜湖: LCG\_RIVER>) 부근에까지 미치고 있는데,
  - <아산호: LCG\_RIVER>는 <남양호: LCG\_RIVER>와 함께 1977년에 국민관광지로 지정되어 유원지로 개발되었으며, 수로(水路)는 낚시터이다.

- <청계천: LCG\_RIVER>의 주요 다리로는 모전교·광통교·장통교·버들다리·두물다리 등이 있다.
- 우리나라에서 제일 규모가 큰 <예당저수지: LCG\_RIVER>에 랜드마크급 음악분수쇼가 생겼다.

#### 6.6.4. 주의 사항

- '4대강'의 경우 '4대강 사업 국가 프로젝트'의 의미를 지닐 경우 TM\_PROJECT로 태깅한다. 줄임말로 사용된 경우라도 국가 프로젝트의 의미라면 TM\_PROJECT로 태깅한다. 한강, 섬진강 등 특정 강의 명칭 없이 단순히 '4대강'만 나왔을 경우 태깅하지 않는다.
- 4대강 사업과 같은 소모성, 낭비성 예산은 철저히 검토해서 전액 삭감해야 되겠다. → <4대강 사업: TMI\_PROJECT>과 같은 소모성, 낭비성 예산은 철저히 검토해서 전액 삭감해야 되겠다.
- '세종보'와 같은 'OO보'의 경우 '댐'과 같은 인공물, 건축물이기 때문에 AF\_BUILDING으로 태깅한다.
- 환경부는 완전 개방 중인 금강 <세종보: AF\_BUILDING>와 <공주보: AF\_BUILDING>를 관측한 결과 멸종위기 야생생물이 출현하는 등 생태계 회복이 확인됐다고 밝혔습니다.

### 6.7. LCG\_OCEAN

#### 6.7.1. 정의 및 설명

- 해양/바다의 명칭

#### 6.7.2. 분석 대상

- 해양
  - 태평양, 인도양 등
- 바다
  - 동해, 남극해, 북극해, 데이비스해, 아문센해, 산호해 등

#### 6.7.3. 예문

- 해양
  - 이어 1521년 세계일주에 나선 마젤란이 <대서양: LCG\_OCEAN>에서 남아메리카 남단의 <마젤란 해협: LCG\_OCEAN>을 지나 <태평양: LCG\_OCEAN>을 거쳐 필리핀에 이르렀다.
  - <인도양: LCG\_OCEAN>에서는 거의 그 중앙부에 <인도양: LCG\_OCEAN> 중앙해령이 남북으로 뻗어 있고,
- 바다

- 한국에서는 조선시대에 <조선해: LCG\_OCEAN>(<朝鮮海: : LCG\_OCEAN>) 또는 <창해: LCG\_OCEAN>(<蒼海: LCG\_OCEAN>)라고 하였으나, 8·15 광복 후부터 <동해: LCG\_OCEAN>라 부르기 시작하였다.
- <산호해: LCG\_OCEAN>(<珊瑚海: : LCG\_OCEAN>)의 일부를 형성하며 서쪽 끝은 <토리스 해협: LCG\_OCEAN>을 거쳐 <아라푸라해: LCG\_OCEAN>(海)로 통한다.

#### 6.7.4. 주의 사항

- 바다의 일부분을 뜻하는 해역도 LCG\_OCEAN로 태깅한다.
  - 바람내리포는 지역주민들이 바람이 불 때 비탈에 바람이 부딪혀 내려와서 "바람내리개"라고 부르고 있다 → <바람내리포: LCG\_OCEAN>는 지역주민들이 바람이 불 때 비탈에 바람이 부딪혀 내려와서 "<바람내리개: LCG\_OCEAN>"라고 부르고 있다.
  - 부산작전기지에서 소말리아 아덴만 해역으로 출항하였습니다. → 부산작전기지에서 소말리아 <아덴만 해역: LCG\_OCEAN>으로 출항하였습니다.

### 6.8. LCG\_BAY

#### 6.8.1. 정의 및 설명

- 반도, 만, 피오르의 명칭

#### 6.8.2. 분석 대상

- 반도
  - 랴오둥반도, 변산반도, 발칸반도, 이베리아반도 등
- 만
  - 도쿄만, 라몬만, 아산만, 울산만 등
- 피오르
  - 송네 피오르드, 밀퍼드 사운드 등

#### 6.8.3. 예문

- 반도
  - 황해연안에 돌출한 <변산반도: LCG\_BAY>는 지괴산지의 대표적인 것으로
  - 따라서 대체로 <발칸 반도: LCG\_BAY>의 기후는 건조하며, 식생(植生)은 특히 남부에서 빈약하다.
  - 포르투갈은 1580년부터 에스파냐의 속령이 되었으며, 1640년에 독립을 회복한 후로는 이들 양국이 <이베리아 반도: LCG\_BAY>를 차지하고 있다.
- 만

- <울산만: LCG\_BAY> 지역은 옛날부터 어업이 성하여
- 충청남도 아산시와 경기도 평택시 사이에 있는 <아산만: LCG\_BAY>은 밀물과 썰물의 차이가 심한 곳인데
- 피오르
  - <송네 피오르: LCG\_BAY>(<Sognefjorden: LCG\_BAY>)는 노르웨이에서 가장 길고 수심이 깊은 피오르이다.

#### 6.8.4. 주의 사항

- '간절곶', '호미곶', '데즈네프 곶' 등 'OO곶'도 LCG\_BAY로 태깅한다.
  - 올해 해돋이는 호미곶에서 보기로 했다. → 올해 해돋이는 <호미곶: LCG\_BAY>에서 보기로 했다.
- '한반도'는 국가와 지형을 모두 의미할 수 있으나 문맥상 지형명으로 사용된 경우만 LCG\_BAY로 태깅한다. 지형명이 아닌 경우는 태깅하지 않는다.
  - 비행기가 한반도 상공을 지나고 있다. → 비행기가 <한반도: LCG\_BAY> 상공을 지나고 있다.
  - '한반도'에는 남한과 북한이 모두 있으므로 국가를 나타내는 LCP\_COUNTRY로 태깅하지 않는다.

### 6.9. LCG\_MOUNTAIN

#### 6.9.1. 정의 및 설명

- 산/산맥, 능선, 광산, 봉(우리)/령, 고개/재의 명칭

#### 6.9.2. 분석 대상

- 산/산맥, 능선
  - 금강산, 라우산, 라웅산, 라키화산, 래밍턴산 등
  - 태백산맥, 차령산맥, 우랄산맥, 공룡능선 등
- 광산
  - 대부광산, 상동광산 등
- 봉(우리)/령, 고개/재
  - 대청봉, 인수봉, 용암봉, 노적봉 등
  - 대관령, 한계령, 미시령 등
  - 노고단, 문경 새재, 박달재 고개 등

#### 6.9.3. 예문

- 산/산맥, 능선
  - 전체 능선의 모양이 마치 닭뿔을 쓴 용의 형상을 닮았다 하여 <계룡산: LCG\_MOUNTAIN>이라고 불린다.
  - 해금강이란 이름은 경치가 <금강산: LCG\_MOUNTAIN>을 닮았다고 하여 남택하(南宅夏)가 지은 것이라고 한다.
  - <태백산맥: LCG\_MOUNTAIN>은 기후풍토상으로도 영동(嶺東) · 영서(嶺西) 양지방의 경계선을 이루며,
  - 또 <차령산맥: LCG\_MOUNTAIN>은 최한월(最寒月) 평균기온 - 3°C의 등온선과 일치하여 한국의 기후 구를 남부의 온대와 북부의 냉대로 크게 구분하는 경계가 되고 있다.
- 광산
  - TV 방송에 나온 이후 안산의 <대부광산: LCG\_MOUNTAIN>은 관광 명소가 되었다.
- 봉(우리)/령, 고개/재
  - 최고봉은 <대청봉: LCG\_MOUNTAIN>이다. <대청봉: LCG\_MOUNTAIN> 남쪽에 <한계령: LCG\_MOUNTIAN>, 북쪽에 <마등령: LCG\_MOUNTIAN>·<미시령: LCG\_MOUNTIAN> 등의 고개가 있다.

#### 6.9.4. 주의 사항

- '꺄똥 고개, 삼년 고개, 바람 고개' 등 정식 명칭이 아닌 명칭은 태깅하지 않는다.
  - 넘어지면 삼 년 내에 죽는다고 알려진 삼년고개에 관한 전설 → 넘어지면 삼 년 내에 죽는다고 알려진 삼년고개에 관한 전설

### 6.10. LCG\_ISLAND

#### 6.10.1. 정의 및 설명

- 섬, 제도의 명칭

#### 6.10.2. 분석 대상

- 섬
  - 대난지도, 대부도, 백령도 등
  - 세빛섬, 서래섬, 나이팅게일섬, 노시베섬, 로드리게스섬 등
- 제도
  - 로스제도, 마스카렌제도 등

#### 6.10.3. 예문

- 섬



- <대난지도: LCG\_ISLAND> 옆의 작은 섬이라 하여 <소난지도: LCG\_ISLAND>라는 이름이 붙었다.
- 서해안에서 제일 큰 섬으로, 큰 언덕처럼 보인다고 하여 <대부도: LCG\_ISLAND>라고 하였으며, 따오기가 흰 날개를 펼치고 공중을 나는 모습처럼 생겼다 하여 <백령도: LCG\_ISLAND>라고 한다.
- 한강 <서래섬: LCG\_ISLAND>에 유채꽃을 볼 목적으로 출발했다.
- 멕시코의 <코즈멜섬: LCG\_ISLAND>에서 스노클링 하고 온 후기
- 제도
  - <올란드 제도: LCG\_ISLAND>는 발트해 북쪽 보트니아만 어귀에 자리잡은 제도이자 핀란드의 자치령이다.

#### 6.10.4. 주의 사항

- (중복 개체명) '제주도'는 섬의 이름이기도 하지만 행정 구역으로 '제주특별자치도'를 의미하기도 하므로 LCP\_PROVINCE로 태깅한다.
  - 영화 '거미숲'을 막 끝낸 송일곤 감독 역시 제주도 옆의 작은 섬 우도로 여행을 떠났다.
    - 영화 '거미숲'을 막 끝냈던 송일곤 감독 역시 <제주도: LCP\_PROVINCE> 옆의 작은 섬 <우도: LCG\_ISLAND>로 여행을 떠났다 .
      - '훗카이도, 오키나와'와 같이 섬이면서 행정 구역인 경우도 LCP\_PROVINCE로 우선 태깅한다. 문맥상 '제주시, 오키나와시'와 같이 문맥상 '시'가 확실한 경우만 LCP\_CITY로 태깅한다.
- (중복 개체명) '코스라에주'와 같이 '섬'이면서 '주'인 경우 LCP\_PROVINCE로 우선 태깅한다.
  - 코스라에주는 현재 전통 농업과 어업이 주를 이룬다. → <코스라에주: LCP\_PROVINCE>는 현재 전통 농업과 어업이 주를 이룬다.
    - 섬 혹은 주, 섬 혹은 시와 같은 멀티 개체명일 경우 문맥상 확실히 'OO섬/도'일 경우만 LCG\_ISLAND로 태깅한다.
- '윗제환초, 아리산호섬'과 같은 환초와 산호섬도 LCG\_ISLAND로 태깅한다.
  - 아리산호섬은 몰디브(Maldives)의 수도 말레(Male)에서 30분 거리에 있으며 카푸 환초(Kaafu Atoll)로부터 45km 서쪽에 위치한다. → <아리산호섬: LCG\_ISLAND>은 몰디브(Maldives)의 수도 말레(Male)에서 30분 거리에 있으며 <카푸 환초: LCG\_ISLAND>(<Kaafu Atoll: LCG\_ISLAND>)로부터 45km 서쪽에 위치한다.
    - 확실하게 특정한 섬을 지칭하는 경우만 LCG\_ISLAND로 태깅한다. '환초, 산호섬' 등과 같은 보통명사는 태깅하지 않는다.

#### 6.11. LCG\_CONTINENT

### 6.11.1. 정의 및 설명

- 대륙의 명칭

### 6.11.2. 분석 대상

- 대륙
  - 아메리카, 아시아, 오세아니아, 유라시아, 유럽, 북아메리카, 남아메리카, 아프리카 등

### 6.11.3. 예문

- 대륙
  - <유럽: LCG\_CONTINENT>은 육반구(陸半球)의 중앙에 위치하고 있어 세계의 다른 대륙에 쉽게 도달할 수 있는 편리한 위치에 놓여 있다.
  - <북아메리카: LCG\_CONTINENT> 대륙은 대체로 멕시코 남부를 가로지르는 테우안테펙 지협(地峽)을 남쪽 한계로 하고 있으며,
  - 대륙으로서 <아프리카: LCG\_CONTINENT>라는 이름으로 불린 것은 16·17세기 네덜란드의 항해자들이 이곳이 독립된 대륙이라는 사실을 알고 난 뒤부터이다.

### 6.11.4. 주의 사항

- '시베리아'와 같이 대륙의 일부분을 지칭하는 특정한 명칭도 LCG\_CONTINENT로 태깅한다.
  - 시베리아의 추위는 상상을 초월한다. → <시베리아: LCG\_CONTINENT>의 추위는 상상을 초월한다.
  - 블라디보스톡 기차역은 시베리아역의 종착역 혹은 발차역으로, 1893년 처음 문을 열었다. → <블라디보스톡 기차역: LC\_OTHERS>은 <시베리아역: LC\_OTHERS>의 종착역 혹은 발차역으로, 1893년 처음 문을 열었다.
    - 문맥상 '시베리아'가 '기차역'을 의미할 경우 LC\_OTHRRES로 태깅하고, '철로'를 의미할 경우 AF\_ROAD로 태깅한다. 문맥으로 판단하기 어려운 경우 LCG\_CONTINENT로 우선 태깅한다.
- '동남아, 북미/남미' 등 대륙의 일부를 지칭하는 명칭이 줄임말로 사용되었을 경우도 LCG\_CONTINENT로 태깅한다.
  - 동남아는 뭐니뭐니해도 리조트 콧 아니겠냐며 → <동남아: LCG\_CONTINENT>는 뭐니뭐니해도 리조트 콧 아니겠냐며
- (중복 개체명) '남극/북극'의 경우 '남극점/북극점'을 나타내면 LC\_OTHERS로 태깅하고 '대륙'을 나타내는 경우 LCG\_CONTINENT로 태깅한다. 문맥으로 판단하기 어려운 경우 LCG\_CONTINENT로 우선 태깅한다.

## 6.12. LC\_SPACE

### 6.12.1. 정의 및 설명

- 항성, 행성의 명칭
- 위성/유성 및 혜성의 명칭
- 별자리, 성운의 명칭

### 6.12.2. 분석 대상

- 항성, 행성
  - 태양, 달, 지구, 수성, 금성, 토성, 명왕성 등
- 위성/유성 및 혜성
  - 거문고자리 유성군, 단주기혜성, 모어하우스혜성, 헬리혜성 등
- 별자리, 성운
  - 미란다, 사자자리, 게자리, 나침반자리, 물병자리, 오리온자리, 고리성운, 견우성, 게성운 등

### 6.12.3. 예문

- 항성, 행성
  - <태양: LC\_SPACE>은 모든 천체 가운데 가장 두드러진 존재이고,
  - 지금까지 <화성: LC\_SPACE>과 <목성: LC\_SPACE> 사이에서 7천여 개의 행성이 발견되었다. ‘<세종: LC\_SPACE>’ 별도 그 중의 하나이며, 지름 5~6km의 작은 행성이다.
  - <지구: LC\_SPACE>에서 가장 멀리 떨어져 있는 <명왕성: LC\_SPACE>(<冥王星: LC\_SPACE>)까지의 거리는 약 60억km, 즉 빛이 5.5 시간 동안에 가는 거리이다.
- 위성/유성 및 혜성
  - 1759년 <헬리혜성: LC\_SPACE>의 회귀(回歸)를 프랑스에서는 처음 관측한 후, 혜성 연구에 전념하여 15개의 새로운 혜성을 포함하여 21개의 혜성을 검출하였다.
  - <타이탄: LC\_SPACE>(<Titan: LC\_SPACE>) 또는 <티탄: LC\_SPACE>은 <토성: LC\_SPACE>의 위성이다. <토성: LC\_SPACE>의 위성 중에서 가장 큰 천체로, <목성: LC\_SPACE>의 <가니메데: LC\_SPACE>에 이어 두 번째로 크다.
- 별자리, 성운
  - <미란다: LC\_SPACE>를 비롯한 <천왕성: LC\_SPACE>의 위성들도 <천왕성: LC\_SPACE>의 자전면과 거의 일치하는 궤도를 가지고 있다.
  - 그 대표적인 예로 <오리온자리: LC\_SPACE> · <말머리성운: LC\_SPACE> · <백조자리: LC\_SPACE> · <북아메리카성운: LC\_SPACE> · <남십자자리: LC\_SPACE> · <석탄 부대: LC\_SPACE>(<coalsack: LC\_SPACE>) 등이 있으나, 은하면에 따른 영역에 많다.

### 6.12.4. 주의 사항

- '슈퍼문, 반달, 보름달' 등도 '달'을 의미하므로 LC\_SPACE로 태깅한다.
  - 보통 바다 수위는 슈퍼문이 나타날 때 가장 높다. → 보통 바다 수위는 <슈퍼문: LC\_SPACE>이 나타날 때 가장 높다.

## 6.13. LC\_OTHERS

### 6.13.1. 정의 및 설명

- 기타 지역/장소의 명칭
- LC 계열의 세부유형이 아닌 기타 장소의 명칭
  - 지하철/기차역, 공항, 버스터미널, IC, 정류장/정거장/주차장, 시장, 마을 등의 구체적인 명칭
  - 농장, 야영장, 체험장, 수련장, 해수욕장 등의 구체적인 명칭
  - 식물원, 수목원, 과수원/농원, 숲, 목장, 정원, 공원 등의 구체적인 명칭
  - 세트장, 광장, 테마파크/놀이 공원 등의 구체적인 명칭
  - 항구/항만/포구, 골짜기, 해안/해변, 온천, 폭포, 계곡, 동굴, 샘터, 약수터 등의 구체적인 명칭
  - 언덕/오름, 나루, 금광, 사구, 발전소 등의 구체적인 명칭
  - 유전, 사막, 오아시스, 갯벌 등의 구체적인 명칭
  - 휴게소, 대피소, 쇼핑/경제/산업/호텔/관광/아파트 단지 등의 구체적인 명칭
  - 특정한 지역/권역을 나타내는 명칭

### 6.13.2. 분석 대상

- 지하철/기차역, 공항, 버스/여객 터미널, IC, 정류장/정거장/주차장, 시장, 마을 등의 구체적인 명칭
  - 서울역, 신도림역, 인천국제공항, 서울남부터미널, 수원IC, 광장시장, 모란시장, 자갈치시장, 남대문 정거장, 서울역 광장, 토성민속마을, 두레미마을, 금당실정보화마을 등
- 농장, 야영장, 체험장, 수련장, 해수욕장 등의 구체적인 명칭
  - 노안청소년야영장, 새마을주말농장, 고현 어촌 체험 마을, 새내역사체험마을, 예송리해수욕장 등
- 식물원, 수목원, 과수원/농원, 숲, 목장, 정원, 공원 등의 구체적인 명칭
  - 다도해 해상국립공원, 계룡산국립공원, 여미지식물원, 국립수목원, 미소정원, 이안숲속, 샘물배농원, 단심농원 등
- 세트장, 광장, 테마파크/놀이 공원 등의 구체적인 명칭
  - 간현유원지, 대성산유원지, 5·18민주광장, 광화문 광장, 송상현 광장, 에버랜드, 디즈니월드, 디즈니랜드, 서울랜드 등

- 항구/항만/포구, 골짜기, 해안/해변, 온천, 폭포, 계곡, 동굴, 샘터, 약수터 등의 구체적인 명칭
  - 부산항, 인천항, 소래포구, 장림포구, 수통골 등
  - 해운대(해안), 용머리 해안, 닭머리 해안, 양남해변, 백암온천, 용정약수터 등
- 언덕/오름, 구릉, 나루, 금광, 사구, 발전소 등의 구체적인 명칭
- 가로구릉, 기아나고지, 나가구릉, 구드레 나루, 미시시피 삼각주, 동아프리카지구대, 태안화력발전소 등
- 유전, 사막, 오아시스, 갯벌 등의 구체적인 명칭
  - 북해 유전, 북자파라나 유전, 사하라 사막, 파라프라 오아시스, 리와 오아시스, 고창 갯벌 등
- 휴게소, 대피소, 쇼핑/경제/산업/호텔/아파트/관광 단지 등의 구체적인 명칭
  - 금강 휴게소, 죽전 휴게소, 장터목대피소, 소청대피소 등
  - 롯데몰, 영통/파주 프리미엄아웃렛, 스타필드, 월스트리트, 천안산업단지, 성서산업단지, 보르시벨트, 주공5단지, 길음뉴타운 8단지, 중문관광단지, 보문관광단지 등
- 특정한 지역/권역을 나타내는 명칭
  - 서해안/동해안/남해안, 서양/동양, 중화권/이슬람권, 동구/서구, 북방/서방/남방/동방, 적도 등

### 6.13.3. 예문

- 지하철/기차역, 공항, 버스/여객 터미널, IC, 정류장/정거장/주차장, 시장, 마을 등의 구체적인 명칭
  - <서울역: LC\_OTHERS>에서 <신도림역: LC\_OTHERS>까지 얼마나 걸리는지 찾아보았다.
  - 연휴 기간에는 <인천국제공항: LC\_OTHERS>과 같은 공항이나 <서울남부터미널: LC\_OTHERS> 같은 터미널에 많은 사람들이 모인다.
  - 성남시는 <모란시장: LC\_OTHERS>의 일부 구역을 재정비하기로 했다.
  - #부산 #주말 #<자갈치 시장: LC\_OTHERS> #물회
  - 축제의 주 행사인 산행은 오전 9시 명석면 <홍지마을 주차장: LC\_OTHERS>에서 모여 준비 운동으로 몸을 풀고 출발하며, 해발 420m 의 <광제봉수대: LC\_OTHERS>에 올라 주변 경관을 감상한 뒤 <홍지마을: LC\_OTHERS>까지 되돌아오는 13구간에서 실시된다.
- 농장, 야영장, 체험장, 수련장, 해수욕장 등의 구체적인 명칭
  - <용연 청소년 야영장: LC\_OTHERS>은 청소년들에게 건전한 야영 활동과 심신 수련 및 호연지기를 키워 주기 위하여 건립했다.
  - <새마을주말농장: LC\_OTHERS>은 광명시 새마을회에서 운영하고 있는 주말농장이다.
  - <한밤실농촌체험마을: LC\_OTHERS>은 2010년, 녹색농촌체험마을로 선정되었다.
  - 울진군에서는 1984년 죽변 <봉평해수욕장: LC\_OTHERS>과 기성 <구산해수욕장: LC\_OTHERS>을 개장한 이래 1997년까지 6개의 해수욕장이 추가로 개장되어 1998년 현

재 총 8개의 해수욕장이 개장되어 있다.

- 식물원, 수목원, 과수원/농원, 정원, 공원 등의 구체적인 명칭
  - 금강이 계룡산을 휘도는 지점에 있는 <금강자연휴양림: LC\_OTHERS>은 강가 팔각정에서 금강과 창벽의 절경이 한눈에 들어온다.
  - <자생식물학습장: LC\_OTHERS>은 우리나라의 산과 들에서 자라고 있는 여러 가지 식물을 모아서 도시 어린이들이 그 생태를 익힐 수 있도록 한 곳이다.
  - 우리 <샘물배농원: LC\_OTHERS>의 배즙은 첨가물 없이 100%를 자랑합니다.
- 세트장, 광장, 테마파크/놀이 공원 등의 구체적인 명칭
  - <에델바이스 스위스테마파크: LC\_OTHERS>는 경기도 가평군 설악면에 위치해 있으며, 스위스의 작은 마을 축제를 주제로 만들어진 곳이다.
  - 주말에 가족과 함께 <용인 대장금파크: LC\_OTHERS>에 놀러갔다.
  - <에버랜드: LC\_OTHERS>의 새로운 놀이공원은 물을 주제로 한 테마파크이다.
- 항구/항만/포구, 골짜기, 해안/해변, 온천, 폭포, 계곡, 동굴, 샘터, 약수터 등의 구체적인 명칭
  - 일몰과 일출을 한 자리에서 볼 수 있어 많은 관광객들이 찾고 있는 충남 서천 <마량 포구: LC\_OTHERS>에서 올해를 마무리하는 풍성한 이벤트가 펼쳐진다.
  - <제주항: LC\_OTHERS>은 최근 국제크루즈선 입항 증가로 국제 관광항만으로서의 역할도 하고 있다.
  - <물만골: LC\_OTHERS>은 수원이 풍부해 골짜기로 물이 많이 흘러내린다 하여 붙은 이름이다.
  - #<해운대: LC\_OTHERS> #주말 여행 #당일치기 여행
  - <동해안: LC\_OTHERS>에 있는 <망상해변: LC\_OTHERS>은 주변에 역도 있고 주말에 바람 쐬기 좋은 곳이다.
  - 외할머니댁 근처에 유명한 <용정약수터: LC\_OTHERS>가 있다.
- 언덕/오름, 구릉, 나루, 금광, 사구, 발전소 등의 구체적인 명칭
  - 부소산성 서쪽으로 백마강가의 <구드레나루: LC\_OTHERS>는 지금은 공원으로 꾸며져 있다.
  - 이번 숙소는 <샛별오름: LC\_OTHERS> 근처로 5분 거리입니다.
  - 유네스코가 기후 및 생물 다양성의 생태계적 가치를 인정하여 제주도를 생물권보전지역으로 지정하였는데 여러 개의 오름 중에서 <거문오름: LC\_OTHERS>이 선정됐다.
  - 그러다가 5,000년 전부터 삼각주의 퇴적환경이 만들어지고, 현재의 <미시시피 삼각주: LC\_OTHERS>가 이루어졌다.
  - 오래전부터 잘 알려진 것은 <동아프리카대지구대: LC\_OTHERS>로, 남단은 모잠비크에서부터 북단은 요르단 근처까지 길이 6,000km에 이르며, 그 내부에서 여러 분지(分枝)로 갈라져 있다.
- 유전, 사막, 오아시스, 갯벌 등의 구체적인 명칭

- 영국 영해와 노르웨이 영해에 걸친 <북해유전: LC\_OTHERS>은 1967년 덴마크 앞바다에서 유전이 발견된 후부터는 석유·천연가스의 자원면에서도 주목받게 되었다.
- <시와 오아시스: LC\_OTHERS>는 이집트의 <사하라 사막: LC\_OTHERS>과 <리비아 사막: LC\_OTHERS>의 사이에 있는 남북 15~25km, 동서 약 160km, 면적 약 300km<sup>2</sup>의 오아시스로 대추야자를 재배한다.
- 고창군에 위치한 갯벌을 <고창 갯벌: LC\_OTHERS>(내만형 갯벌)이라고 합니다.
- 휴게소, 대피소, 쇼핑/경제/산업/호텔/아파트/관광 단지 등의 구체적인 명칭
- 봉정암에서 <소청대피소: LC\_OTHERS>를 거쳐 대청봉까지
- 드디어 <죽전 휴게소: LC\_OTHERS>, 곧 서울이다.
- 간만에 쇼핑물 움 #<롯데프리미엄아울렛: LC\_OTHERS> #평일의 여유 #혼자 쇼핑
- 드디어 와 보는 <스타필드: LC\_OTHERS>, 진짜 넓다. <수원 롯데몰: LC\_OTHERS> 같다.
- <구로디지털단지: LC\_OTHERS>와 <가산디지털단지: LC\_OTHERS>가 구성되어 있으며, 옛날에는 <구로공단: LC\_OTHERS>으로 불리던 산업지역이다.
- <보르시 벨트: LC\_OTHERS>(<Borscht Belt: LC\_OTHERS>)는 <미국: LCP\_COUNTRY> <뉴욕주: LCP\_PROVINCE> <캐츠킬 산맥: LCG\_MOUNTAIN>에 있는 유대인들의 호텔 단지를 가리키는 용어이다.
- <길음뉴타운: LC\_OTHERS>이랑 <중계5단지: LC\_OTHERS>랑 가격이 많이 올라서
- <중문관광단지: LC\_OTHERS>에서 즐길 거리 찾기
- 특정한 지역/권역을 나타내는 명칭
  - 북태평양 고기압이 팽창하면서 오랜 시간 <남해안: LC\_OTHERS>에 비를 뿌리고 있습니다.
  - 선진국과도 통하고 후진국과도 통하고 <서양: LC\_OTHERS>과도 통하고 <동양: LC\_OTHERS>과도 통할 수 있는 그런 여러 가지 중재적인 성격을 가진 <한국: LCP\_COUNTRY>이라는 나라의 특성 때문에 이것이 <한국: LCP\_COUNTRY>에서 열렸다.
  - 대만 역시 <중화권: LC\_OTHERS>이라는 느낌이 들었다.
  - 세계 역사를 주도했던 <유라시아: LCG\_CONTINENT>는 해양을 통한 <서구: LC\_OTHERS>의 침탈과 지배 하에 수 세기 동안 놓여 있었다.

#### 6.13.4. 주의 사항

- '효녀 마을, 장수 마을'과 같이 정식 명칭이 아니거나, '강, 바다, 대륙붕', '간석지, 우발레, 주상해분', '슬로프, 마을/시장, 검문소/무기제작소/군기창고/군사보호구역/군사훈련장', '수산(물)/민속 시장, 해안/해변, 놀이동산, 쇼핑/아파트 단지' 등 특정 명칭 없이 보통명사로 사용한 경우 태깅하지 않는다.
  - 이집트를 흐르는 나일강은 정기적으로 범람하는 강이다. → <이집트: LCP\_COUNTRY>를 흐르는 <나일강: LCG\_LIVER>은 정기적으로 범람하는 강이다.
  - 여름 휴가는 산보다 바다가 낫지. 올해는 태안 반도? → 여름 휴가는 산보다 바다가 낫

- 지. 올해는 <태안 반도: LCG\_BAY>?
- 이 근처에서 제일 가까운 놀이동산이나 동물원이 어디야? → 이 근처에서 제일 가까운 놀이동산이나 동물원이 어디야?
  - '이스턴 슬로프, 루키 슬로프'와 같은 슬로프명은 LC\_OTHERS로 태깅한다.
    - 12월 10일 화요일 이스턴슬로프가 추가 오픈됩니다. → 12월 10일 화요일 <이스턴슬로프: LC\_OTHERS>가 추가 오픈됩니다.
  - 검문소/무기제작소/군기창고/군사보호구역/군사훈련장의 경우 'OO검문소'와 같은 특정한 명칭인 경우 LC\_OTHERS로 태깅한다.
    - 강모 씨가 승용차 운전 중 고성군 배둔검문소 앞 도로에 떨어져 있던 청구조물과 부딪혀 승용차가 파손되었다. → 강모씨가 승용차 운전 중 <고성군: LCP\_COUNTY> <배둔검문소: LC\_OTHERS> 앞 도로에 떨어져 있던 청구조물과 부딪혀 승용차가 파손되었다.
  - '게토' 등의 특정 구역을 가리키는 것은 LC\_OTHERS로 태깅한다.
    - 6세기 경 콘스탄티노플에는 서부 유럽의 도시들보다 한참 전에 게토가 있었다. → 6세기 경 <콘스탄티노플: LCP\_CAPITALCITY>에는 <서부: TM\_DIRECTION> <유럽: LCG\_CONTINENT>의 도시들보다 한참 전에 <게토: LC\_OTHERS>가 있었다.
  - '올드페이스풀 간헐천' 등과 같은 OO간헐천도 온천이므로 LC\_OTHERS로 태깅한다.
    - 미국 옐로스톤 국립공원의 옐로스톤 칼데라 안에 있는 올드페이스풀 간헐천은 40분에서 80분 간격으로 분출한다. → 미국 <옐로스톤 국립공원: LC\_OTHERS>의 옐로스톤 칼데라 안에 있는 <올드페이스풀 간헐천: LC\_OTHERS>은 40분에서 80분 간격으로 분출한다.
  - '울산바위, 갯바위'와 같은 OO바위는 특정한 공간의 의미가 아니므로 태깅하지 않는다.
    - 속초를 병풍처럼 감싸고 있는 설악산을 바라보면 전면에 보이는 바위가 바로 울산바위다.
  - '통일전망대, 앞산공원전망대'와 같은 'OO전망대'의 경우 인공물, 건축물이기 때문에 AF\_BUILDING으로 태깅한다.
    - <통일전망대: AF\_BUILDING>는 다소 아득하게 느껴졌던 분단의 현실을 직시할 수 있는 곳이다.
  - '서면, 광안리, 압구정로데오'와 같이 정확한 행정 구역 명칭이 아니며 지하철역 이름으로 그 일대를 일컫는 경우 LC\_OTHERS로 태깅한다.
    - 지난 주말에는 대학 동기들과 시간을 맞춰 <서면: LC\_OTHERS> 맛집에 다녀왔는데요.
  - (중복 개체명) '전주 객사'와 같이 구체적인 유적지, 마을, 터, 성지, 전승지 등이 문화재로 등록되어 있으면 AF\_CULTURAL\_ASSET로 태깅하고, 문화재가 아니면 LC\_OTHERS로 태깅한다.
    - 토성민속마을과는 달리 문화재로 지정된 아산외암민속마을은 주의 사항이 더 많다. → <토성민속마을: LC\_OTHERS>과는 달리 문화재로 지정된 <아산외암민속마을: AF\_CULTURAL\_ASSET>은 주의 사항이 더 많다.
  - (중복 개체명) '경포대'는 문맥에 따라 해변, 건물 등으로 모두 태깅 가능하다. 이때 문맥과



적정술어로 판정할 수 있다. '붐비다, 놀러 가다' 등의 술어와 함께 사용되면 LC\_OTHERS, '세우다, 만들다' 등의 술어와 사용되면 AF\_BUILDING으로 태깅한다. 이때 술어를 참고하되 술어의 다의성에 주의하여 태깅하며, 문맥으로 판단이 어려울 경우 LC\_OTHERS로 우선 태깅한다.

- 동해안의 대표적인 해변인 경포대 일출은 사실 어느 지점에서 봐도 동일해요. → 동해안의 대표적인 해변인 <경포대: LC\_OTHERS> 일출은 사실 어느 지점에서 봐도 동일해요.
- 살짝 보이는 누각, 강릉 경포대를 한눈에 담아 봅니다. 처마부분이 정말 화려하지요? → 살짝 보이는 누각, 강릉 <경포대: AF\_BUILDING>를 한눈에 담아 봅니다. 처마 부분이 정말 화려하지요?
- (중복 개체명) '해운대'와 같은 경우 해수욕장/해변, 행정구역, 대로, 지하철역을 모두 나타낼 수 있다. 해수욕장/해변일 경우 LC\_OTHERS로 태깅하고, 주소로 사용되어 '해운대구'를 의미하면 LCP\_COUNTY, '해운대로'면 AF\_ROAD, '해운대역'을 의미하면 LC\_OTHERS로 태깅한다. 문맥으로 판단하기 어려운 경우 LC\_OTHERS로 우선 태깅한다.
  - 부산 해운대 가 볼 만한 곳 10 → <부산: LCP\_CITY> <해운대: LC\_OTHERS> 가 볼 만한 곳 10
  - 부산 해운대 우동의 한 아파트에서 불이 나 주민 수십 명이 대피하고 1명이 다쳤다. → <부산: LCP\_CITY> <해운대: LCP\_COUNTY> <우동: LCP\_COUNTY>의 한 아파트에서 불이 나 주민 수십 명이 대피하고 1명이 다쳤다.
- (중복 개체명) '서면롯데백화점, 현대백화점 판교'와 같이 특정한 'OO백화점'이나 '홍대/이대' 등 특정한 'OO대(학교)'와 같은 OG계열이 '~에서 만나다/구입하다/팔다/사다' 등과 함께 사용되어 문맥상 '기관'을 의미하지 않고 '장소/위치'를 의미하는 경우 LC\_OTHERS로 태깅한다. 이때 술어를 참고하되 술어의 다의성에 주의하여 태깅하며, 문맥으로 판단이 어려운 경우 OGG\_ECONOMY나 OGG\_EDUCATION 등 개체명의 본유적 의미를 우선 고려하여 태깅한다.
  - 어제 홍대에서 산 핑크색 겨울 니트 → 어제 <홍대: LC\_OTHERS>에서 산 핑크색 겨울 니트
  - 서강대 근처에 있는 카페에서 → <서강대: LC\_OTHERS> 근처에 있는 카페에서
  - 그는 광양제철고-중앙대를 거쳐 안산 그리너스에서 프로에 데뷔했다. → 그는 광양제철고-<중앙대: OGG\_EDUCATION>를 거쳐 안산 그리너스에서 프로에 데뷔했다.
  - 제6회 집현전 학술대회는 세종문화회관 예인홀에서 열린다. → 제6회 집현전 학술대회는 <세종문화회관: AF\_BUILDING> 예인홀에서 열린다.
    - '기관'이 아닌 '건축물'의 개념으로 사용되었다고 판단되는 경우, OG가 아닌 AF로 태깅한다. '~를 운영하다, ~를 경영하다, ~이 망하다'와 같은 서술어와 어울리면 OGG로 태깅하고 '~이 열린다'와 같은 서술어와 어울리면 AF\_BUILDING으로 태깅한다.
- '현대백화점 VIP룸', '세종문화회관 예인홀', '경기대학교 체육관'과 같은 경우 '현대백화점'이나 '세종문화회관', '경기대학교'만 태깅하고, 백화점/회관/건물/대학교 안의 장소인 'OO룸,

OO관, OO홀' 등은 태깅하지 않는다.

- '목욕탕, 노래방, PC방, 찜질방' 등도 LC계열이 아니므로 LC\_OTHERS로 태깅하지 않는다. 'OO목욕탕, OO노래방, OOPC방, OO찜질방' 등 특정한 명칭이 있는 경우만 OGG\_ECONOMY로 태깅한다.
- 'OO꽃밭'인 경우 'OO꽃밭'의 OO만 내부 태깅하고 '꽃밭'은 태깅하지 않는다. LC\_OTHERS로 태깅하지 않는다.
  - 양귀비 꽃밭 → <양귀비: PT\_FLOWER> 꽃밭
- 'OO밭'인 경우, 표기 유형은 같으나 사전 등재 여부가 다를 수 있다. 가령 '감자밭/배추밭/고구마밭' 등은 사전에 등재되어 있으나 '딸기밭/수박밭'은 등재되어 있지 않다. 이럴 경우 모두 동일하게 처리하여 내부 태깅하지 않는다.
  - 감자밭 → 감자밭, 딸기밭 → 딸기밭

## 7. CIVILIZATION

---

CIVILIZATION은 문명/문화와 관련된 용어들로 구성되어 있다. 문명/문화, 민족/종족, 언어, 제도, 법/법률, 조세, 연금/기금/자금, 통화, 예술, 스포츠 종목, 스포츠 포지션, 스포츠 용품, 상/훈장, 가족/친족 관계, 직업, 직위/직책/직급, 음식/음료의 명칭, 음식 유형, 섬유/의복, 건축 양식과 관련된 용어가 이에 해당된다. CIVILIZATION 세부 범주에 해당하지 않는 것들은 태깅 대상에서 배제한다.

### 7.1. CV\_CULTURE

#### 7.1.1. 정의 및 설명

- 문명
- 문화
- 혁명/사회 변혁

#### 7.1.2. 분석 대상

- 문명
  - 인더스문명, 황허문명, 아즈텍문명, 안데스문명, 메소포타미아문명
- 문화
  - 거석문화, 날문화, 도식문화
- 혁명/사회 변혁

- 1차산업혁명, 2차산업혁명, 3차산업혁명, 4차산업혁명, 르네상스

### 7.1.3. 예문

- **문명**
  - <인더스문명: CV\_CULTURE>의 대표적 유적은 당시의 2대 도시였던 하라파와 모헨조다로인데,
  - <인더스문명: CV\_CULTURE>을 남긴 사람들의 인종적 구성도 복잡하여, 고대의 모헨조다로에서 발굴된 14개의 인골(人骨)의 예형(例型) 중 3 예형은 원(原)오스트랄로이드, 6 예형은 지중해 인종, 1 예형은 몽골 인종, 4 예형은 알프스 인종이었다.
  - 은나라의 청동기는 매우 특이한 것으로서 <황허문명: CV\_CULTURE>의 상징이다.
  - 이처럼 발달한 <아스테크문명: CV\_CULTURE>은 1520년 갑자기 나타난 에스파냐의 페르난도 코르테스가 거느린 수백 명의 군대의 침입으로 수개월에 걸친 외교교섭과 격렬한 전투 끝에 멸망하였다.
  - <안데스 문명: CV\_CULTURE>의 오랜 역사를 통하여 문자(文字)는 끝내 발명되지 않았다.
  - <메소포타미아 문명: CV\_CULTURE>은 폐쇄적인 <이집트 문명: CV\_CULTURE>과는 매우 대조적이었다.
- **문화**
  - 입석은 선사시대 <거석문화: CV\_CULTURE>의 한 형태로 지역 간의 경계를 나타내거나 액운을 막아 주는 신앙의 대상으로 세워졌다.
  - <암리: CV\_CULTURE>(<Amri: CV\_CULTURE>) · <날문화: CV\_CULTURE>라 병칭되는 바와 같이 <암리문화: CV\_CULTURE>와 많은 연관이 있는데, 분포범위와 가지무늬토기(彩文土器)의 양식이 약간 다를 뿐이다.
  - 이글루릭 에스키모 지역에서 <도싯문화: CV\_CULTURE>에의 이행과정에 대한 연구가 진행되고 있다.
- **혁명/사회 변혁**
  - 오늘은 <4차 산업혁명: CV\_CULTURE>과 한국국토정보공사와의 연계에 대해서 알아보려고 합니다.
  - 독일은 풍부한 석탄과 발전된 철강산업을 바탕으로 최신 기술을 적용한 공장을 속속 만들어 내면서 <2차 산업혁명: CV\_CULTURE>을 선도할 수 있었다.

## 7.2. CV\_TRIBE

### 7.2.1. 정의 및 설명

- 민족 및 종족

- 국가를 구성하는 국민의 지칭

### 7.2.2. 분석 대상

- 민족 및 종족
  - 이라크족, 유대인, 조선족, 마리오족, 아메리카인디언, 아리아인, 고산족, 집시, 보헤미안, 유목민
- 국가를 구성하는 국민의 지칭
  - 한국인, 미국인, 중국인, 일본인

### 7.2.3. 예문

- 민족 및 종족
  - <이라크족: CV\_TRIBE>의 쿰바(Kumba)제도처럼 남녀 모두에게 배우자 이외의 사람과의 성관계를 허용하는 '성관계 범위'가 설정된 사회도 있다.
  - <마오리족: CV\_TRIBE>은 아마도 동부 폴리네시아의 소시에테제도에서 쿡제도를 경유하여 몇 차례에 걸쳐 건너왔으며,
  - 인도에 침입한 <아리아인: CV\_TRIBE>은 펀자브 지방에 들어가 원주민을 정복하고 혼혈을 장려하였다.
  - <고산족: CV\_TRIBE>의 원주지(原住地)나 타이완으로 이주해 온 연대가 종족에 따라 상이하므로 문화수준도 종족 간에 격차가 심하다.
- 국가를 구성하는 국민의 지칭
  - 바빌론유수기를 거쳐 고향으로 돌아온 <이스라엘인: CV\_TRIBE>을 모두 <유대인: CV\_TRIBE>이라 불렀고,
  - 현재 독립운동가의 유족들을 포함한 약 200만 정도로 추산되는 <조선족: CV\_TRIBE>은 국적이 <중국인: CV\_TRIBE>이므로 모국에의 귀환이 성사되기 어려운 현 상황에서는
  - 뒤에 본래의 <인도인: CV\_TRIBE>과 구별하기 위하여 '아메리카의 <인도인: CV\_TRIBE>', 즉 <아메리카인디언: CV\_TRIBE>이라 부르게 되었다.

### 7.2.4. 주의 사항

- (고인류를 분류한 학명) 고인류를 분류한 학명인 '호모사피엔스', '호모사피엔스사피엔스', '호모에렉투스', '네안데르탈인' 등도 CV\_TRIBE로 태깅한다.
- (-계) '계통'의 뜻을 더하는 접미사가 붙은 한국계, 유대인계, 히스패닉계 등도 CV\_TRIBE로 태깅한다.
- (인종) 인류를 지역과 신체적 특성에 따라 구분한 종류도 CV\_TRIBE로 태깅한다.
  - 아프리카인, 아메리카인, 아시아인
  - 흑인, 백인, 황인

## 7.3. CV\_LANGUAGE

### 7.3.1. 정의 및 설명

- 언어의 하위 유형

### 7.3.2. 분석 대상

- 언어의 하위 유형
  - 한국어, 영어, 힌디어, 프랑스어, 이탈리아어, 산스크리트, 라틴어, 몽골어, 아카드어, 말레이어, 스와힐리어

### 7.3.3. 예문

- 언어의 하위 유형
  - 이와 같이 <라틴어: CV\_LANGUAGE>는 학문과 종교의 언어로서 근대 서유럽의 여러 언어에 깊숙이 침투하여 그들 언어에 많은 어휘를 제공하였다.
  - 현대 <몽골어: CV\_LANGUAGE>는 동서의 대방언군(大方言群)으로 나뉘고 그 밖에 약간 고립된 방언이 있으나, 지역적 언어 차는 별로 심하지 않다.
  - 수메르의 설형문자에서 발전하여 설형문자 형태로 표기하는 <아카드어: CV\_LANGUAGE>는 약 600개 낱말과 음절 기호가 있다.
  - <말레이어: CV\_LANGUAGE>는 공통어로서 발달했기 때문에 문법에서 관용도의 폭이 상당히 넓고 또 실제 발음이나 어휘에도 지방의 차이가 심하다.
  - <스와힐리어: CV\_LANGUAGE>의 모든 음절은 개음절(開音節)이다. 어순은 영어의 어순과 비슷하며 동사의 인칭변화(人稱變化)도 있다.

### 7.3.4. 주의 사항

- (외국어 등) '외국어' 등의 보통명사는 CV\_LANGUAGE로 태깅하지 않는다. 개체명으로 보기 어렵다.
  - 외국어, 모국어, 표준어, 방언, 수화, 수어 등
- (방언) '방언', 그 하위 부류로 볼 수 있는 '동남 방언' 등은 모두 '한국어, 영어' 등과 같은 계열로 볼 수 없으므로 태깅 대상이 아니다.
- 단, 다음과 같은 복합어 구성은 전체를 CV\_LANGUAGE로 태깅한다. 한국 수어, 한국 수화 언어의 경우 전체를 CV\_LANGUAGE로 태깅한다. '한국 수어, 한국 수화 언어'는 모두 우리 말샘에 등재되어 있다. 이에 근거하여 '프랑스 수어, 일본 수어' 등도 CV\_LANGUAGE로 태깅한다.
  - 프랑스 수어 → <프랑스 수어: CV\_LANGUAGE>
  - 한국 수화 언어 → <한국 수화 언어: CV\_LANGUAGE>

- ('-어'가 빠진 경우) '일본 버전 책'의 경우 문맥에서 '일본'이 '일본어'를 가리키는 것임을 명확히 알 수 있다면 CV\_LANGUAGE로 태깅하고, 알 수 없다면 LCP\_COUNTRY로 태깅한다.
  - <한국: LCP\_COUNTRY> 버전 '보보경심'은 원작 소설과 많이 다르지만, <중국: LCP\_COUNTRY> 버전 '보보경심'은 소설과 거의 비슷하다.
  - <영어: CV\_LANGUAGE> 버전으로 먼저 출간된 이 한국어 교재는 현재 <중국: CV\_LANGUAGE> 버전과 <일본: CV\_LANGUAGE> 버전의 번역 작업이 진행 중이다.
    - '일본 버전 책'처럼 일본어로 된 책인지, 아니면 소설 등이 일본의 문화에 맞게 지어진 책인지 알 수 없을 경우, '<일본: LCP\_COUNTRY> 버전 책'과 같이 태깅한다.

## 7.4. CV\_POLICY

### 7.4.1. 정의 및 설명

- 제도 및 정책의 명칭

### 7.4.2. 분석 대상

- 제도 및 정책의 명칭
  - 카스트, 쓰레기종량제, 직업공무원제도, 연금제도, 고시제도
  - 가부장제, 진대법, 노예제, 교과담임제, 최저임금제도
  - 교과담임제, 최저임금제도, 섯다운제도

### 7.4.3. 예문

- <가부장제: CV\_POLICY>의 가족 형태에서는 가족성원이 세습적 규칙에 따라 지명된 개인의 지배를 받는데, 대개는 장남이 세습적으로 가장의 지위와 재산을 계승하여 안으로는 가족을 통솔하고 밖으로는 가족을 대표한다.
- 10월에 환납(還納)하는 것을 상식(常式)으로 시행하도록 하였는데, 이것이 한국에서의 <진대법: CV\_POLICY> 실시의 최초의 기록이다.
- <노예제: CV\_POLICY> 안에는 각양각색의 사회적 신분을 가진 자가 포함된다.
- 현재 초등학교는 <학급담임제: CV\_POLICY>를, 중학교 이상은 <교과담임제: CV\_POLICY>를 채택하고 있다.
- <최저임금제도: CV\_POLICY>를 통한 국가의 강제에 의한 임금액의 보호는 노사간에 실질적인 교섭 평등관계가 유지되고 있지 않은 사업장의 근로자들을 위해서는 절실히 요구된다.

### 7.4.4. 주의 사항

- 제가회의, 화백회의, 정사암회의 등은 CV\_POLICY로 태깅하지 않는다. 개체명으로 보기 어렵다.
- 국가 사업은 TMI\_PROJECT로 태깅한다. 정책, 제도에 한해서 CV\_POLICY로 태깅한다. 따라서 '화폐정리사업'은 CV\_POLICY가 아닌 TMI\_PROJECT로 태깅한다.
- 교육복지 정책, 결혼 제도는 세부적인 제도, 정책으로 볼 수 없기 때문에 태깅하지 않는다. 아동수당, 주택보조금, 무상급식 등은 각각 특정한 제도, 정책이기 때문에 CV\_POLICY로 태깅한다.
- 계엄령, 비상계엄, 경비계엄 등과 같이 실제 시행이 되는 정책 용어인 경우는 법률 용어이지만 CV\_POLICY로 태깅한다.

## 7.5. CV\_LAW

### 7.5.1. 정의 및 설명

- 법/법률 명칭
- 법적 권리
- 범죄 명칭
- 형벌 명칭
- 칙령
- 협정, 협약의 명칭

### 7.5.2. 분석 대상

- 법/법률 명칭
  - 국제연합헌장, 민법, 형법, 함무라비법전, 관습법 등
  - 헌법, 국회법, 법원조직법 등
- 법적 권리
  - 피선거권, 참정권, 선거권, 공민권 등
  - 신분권, 친권, 부양청구권, 상속권, 어업권, 상표권, 저작권(소프트웨어 라이선스) 등
- 범죄 명칭
  - 간통, 살인죄, 사체유기죄, 폭행죄, 공무집행방해죄, 명예훼손죄, 사기죄 등
- 형벌 명칭
  - 화형, 태형, 구금 등
  - 구류, 벌금, 몰수, 교수형, 총살형 등
- 칙령
  - 밀라노 칙령, 베를린 칙령 등
- 협정, 협약의 명칭

- 국제주석협정, 나소협정, 베를린협정 등
- 미쓰야협약, 런던협약, 간도협약 등

### 7.5.3. 예문

- 법/법률 명칭
  - 실질적 의미의 <헌법: CV\_LAW>에는 형식적 의미의 <헌법: CV\_LAW>뿐 아니라 <정부조직법: CV\_LAW> · <국회법: CV\_LAW> · <법원 조직법: CV\_LAW> · <정당법: CV\_LAW> · <선거법: CV\_LAW> 등의 법률과
- 법적 권리
  - 대의원 자격 요건은, 국회의원 <피선거권: CV\_LAW>이 있는 30 세 이상의 자로서 평화통일을 위해 주권을 행사할 수 있는 인물이어야 했다.
  - 법원의 판결 또는 다른 법률에 의하여 <피선거권: CV\_LAW>이 정지되거나 상실된 자는 <피선거권: CV\_LAW>이 없다.
  - <참정권: CV\_LAW>(<參政權: CV\_LAW>)도 광범위하게 보장되며,
  - 1회원이 가질 출자좌수는 총출자좌수의 100분의 15를 초과할 수 없다. 회원은 출자좌수에 관계없이 평등한 <의결권: CV\_LAW>과 <선거권: CV\_LAW>을 가진다.
  - 수형(受刑) 중인 사람이나 선거법을 위반한 자에 대해서는 <선거권: CV\_LAW> · <피선거권: CV\_LAW>이 박탈되며, 이를 <공민권: CV\_LAW>의 정지 · 박탈이라고 한다.
  - <신분권: CV\_LAW>에는 <친족권: CV\_LAW>과 <상속권: CV\_LAW>이 있는데, <친권: CV\_LAW>, <후견권: CV\_LAW>, 배우자에 대한 권리, <부양 청구권: CV\_LAW> 등은 <친족권: CV\_LAW>에 속하며, <재산상속권: CV\_LAW>은 <상속권: CV\_LAW>에 속한다.
  - <어업권: CV\_LAW>은 법적 성질이 <물권: CV\_LAW>임에도 불구하고 원칙적으로 이를 이전 · 분할 또는 변경할 수 없다.
  - <상표권: CV\_LAW>은 <재산권: CV\_LAW>의 일종으로서 <특허권: CV\_LAW> 등과 같이 담보에 제공될 수 있으며, 지정상품의 영업과 함께 이전할 수도 있다.
- 범죄 명칭
  - 사체를 다른 곳으로 옮겨 놓은 경우, 또는 그 죄적을 은폐하기 위하여 사체를 은닉한 경우에는 <살인죄: CV\_LAW>와 <사체유기죄: CV\_LAW> 또는 <사체은닉죄: CV\_LAW>와의 경합범(競合犯)이 된다.
  - <폭행죄: CV\_LAW>의 폭행은 사람의 신체에 대한 일체의 불법적인 유형력의 행사를 포함하게 된다.
  - <공무집행방해죄: CV\_LAW>는 공무를 집행하는 공무원에 대하여 폭행 또는 협박하여 직무 집행을 하지 못하도록 하는 경우와
  - 명예훼손의 행위(사실 또는 허위의 사실을 적시한 행위)와 모욕행위가 동시에 행해졌을 때에는 모욕행위는 <명예훼손죄: CV\_LAW>에 흡수되어 <명예훼손죄: CV\_LAW>만이 성립한다.



- 결혼을 가장하고 상대방을 기만하여 결혼과 패물 비용 등의 명목으로 금품을 사취하는 이른바 혼인사기는 <사기죄: CV\_LAW>가 된다.
- 형벌 명칭
  - 한국에서는 <화형: CV\_LAW>을 제도화하거나 관습화되었던 사례는 찾아볼 수 없으나
  - 사형수는 모두 <참형: CV\_LAW>으로 집행하였다.
  - <구금: CV\_LAW>이 비교적 장기에 걸치는 강제처분인 데 반해 <구류: CV\_LAW>는 가벼운 범죄에 대하여 형법상 다른 형과 함께 선택형으로 규정하고 있으며,
  - <벌금: CV\_LAW>은 판결 확정일로부터 30일 이내에 납입하여야 하며,
  - 행위자에게 유죄의 재판을 하지 아니할 때에도 <몰수: CV\_LAW>의 요건이 있는 때에는 <몰수: CV\_LAW>만을 선고할 수 있다.
  - 한국에서는 원칙적으로 <교수형: CV\_LAW>이 사형집행 방법으로 되어 있고, 군인은 원칙적으로 <총살형: CV\_LAW>이 집행된다.

#### 7.5.4. 주의 사항

- (법률용어) 헌법, 국회, 대통령을 거쳐 성립된 국법이 아닌 아래와 같은 법률용어는 CV\_LAW로 태깅하지 않는다.
  - 재판 절차와 관련한 용어: 상고, 상소, 항소심
  - 제도 또는 명령과 관련한 용어: 탄핵, 소추, 탄핵소추, 탄핵심판, 무죄판결, 긴급명령, 긴급재정경제처분
  - 특정한 법안이나 용어가 아닌 경우: 법률안, 예산안, 수정안 등
- 계엄령, 비상계엄, 경비계엄 등과 같이 실제 시행이 되는 정책 용어인 경우는 법률 용어이지만 CV\_POLICY로 태깅한다.
- (법률용어) 감사 활동, 특정 처분이나 원칙을 나타내는 법률용어는 CV\_LAW로 태깅한다.
  - 국정감사, 불기소처분, 일사부재의
- (명령 관련) 특정 정책이나 행정 명령은 CV\_LAW로 태깅한다.
- 방곡령, 대사령 → CV\_LAW
- (일상적 명칭의 법률) 민식이법, 김영란법 등 새로 생겨난 법안은 CV\_LAW로 태깅한다.
- (~법) '대동법'은 조세제도이므로 CV\_POLICY로도 볼 수 있지만, ~법으로 끝나는 경우에는 CV\_LAW로 태깅한다.
  - 벌금 200만 원, 징역 12개월과 같은 처벌(형벌)은 각각 분리하여 태깅한다. 단, 벌금 OO형, 징역 OO형, 집행유예 OO형은 통합하여 태깅한다.
  - 벌금 200만 원 → <벌금: CV\_LAW> <200만 원: QT\_PRICE>
  - 징역 12개월 → <징역: CV\_LAW> <12개월: DT\_DURATION>
  - 징역 1년, 집유 6월 → <징역: CV\_LAW> <1년: DT\_DURATION>, <집유: CV\_LAW> <6월: DT\_DURATION>
  - 징역 3년형, 집행유예 1년형 → <징역 3년형: CV\_LAW>, <집행유예 1년형: CV\_LAW>

- 법률 조항의 경우도 각각 분리하여 태깅한다.
  - OO법 제3조 제 4항 → <OO법: CV\_LAW>, <제3조: QT\_ORDER> <제 4항: QT\_ORDER>
- '벌금'은 법적 효력이 있는(벌금형 선고) 경우에만 태깅하며, 한 사회나 공동체에서 규칙으로 정한 벌금의 경우에는 태깅하지 않는다.
- (중복 개체명) '~조약', '~협정'은 문맥에 준하여 태깅한다.
  - 1905년 9월 5일, 미국의 중재로 일본과 러시아가 미국에서 포츠머스 조약을 체결했다.  
→ 1905년 9월 5일, 미국의 중재로 일본과 러시아가 미국에서 <포츠머스 조약: EV\_OTHERS>을 체결했다.
  - 남북 평화협정 맺고 북미수교로 나아가야. → <남북 평화협정: EV\_OTHERS> 맺고 북미수교로 나아가야.
  - <포츠머스 조약: CV\_LAW>에는 러시아가 조선으로부터 철수하고 모든 영향력을 내려놓겠다는 내용이 쓰여 있다.
  - 단, 문맥으로 판단하기 어려울 경우, EV\_OTHERS로 태깅한다.

## 7.6. CV\_CURRENCY

### 7.6.1. 정의 및 설명

- 통화 명칭

### 7.6.2. 분석 대상

- 통화 명칭
  - 달러, 엔, 유로, 루피, 위안, 원

### 7.6.3. 예문

- <유로: CV\_CURRENCY>는 7종의 지폐와 8종의 동전으로 구성되며 유로의 제작·발행은 각 나라가 독자적으로 실시한다.
- <루피: CV\_CURRENCY>는 '은'을 나타내는 산스크리트 '루피아'에서 유래한다.
- <위안: CV\_CURRENCY>의 대외적 환율은 1950년 7월 중국인민은행이 환율을 공표함으로써 전국적으로 통일된 환율이 실현되었고,

## 7.7. CV\_TAX

### 7.7.1. 정의 및 설명

- 조세 명칭

### 7.7.2. 분석 대상

- 조세 명칭
  - 소득공제, 면허세, 관세
  - 부가가치세, 근로소득세, 재산세, 취득세, 부가세

### 7.7.3. 예문

- <부가가치세: CV\_TAX>는 모든 재화 또는 용역의 소비행위에 대하여 부과되는 일반소비세이며,
- 일용근로자의 근로소득에 대한 <근로소득세: CV\_TAX>는 원천징수에 의하여 과세가 종결된다.
- <재산세: CV\_TAX>는 건축물·선박 및 항공기를 과세물건(課稅物件)으로 하며(동법 제180조, 제181조),
- <취득세: CV\_TAX>는 토지·건축물·차량·기계장비·입목(立木)·항공기·선박·광업권·어업권·골프회원권·콘도미니엄회원권·종합체육시설 이용회원권 등의 자산의 취득에 대하여 그 취득자에게 부과한다(지방세법 제105조).
- 공공단체(公共團體)가 증징(增徵)하는 세액에 대하여 <부가세: CV\_TAX>를 부과하도록 하였었다.

## 7.8. CV\_FUNDS

### 7.8.1. 정의 및 설명

- 연금, 기금, 자금, 펀드 명칭

### 7.8.2. 분석 대상

- 연금, 기금, 자금, 펀드 명칭
  - 거치연금, 문예진흥기금, 공적자금, 퇴직연금
  - 노령연금, 개인연금, 통화안정기금, 방송발전기금, 공적자금

### 7.8.3. 예문

- 가입 기간이 20년 이상인 가입자 또는 가입자였던 자가 60세(대통령령이 정하는 특수직종 근로자는 55 세)에 달한 때에는 그때부터 그가 생존하는 동안 <노령연금: CV\_FUNDS>을 지급받는다.
- <개인연금: CV\_FUNDS>은 정부의 정책지원에 의해 저축기간 동안 연간 최고 72만 원을 한도로 불입액의 40%가 소득공제되며, 이자소득세도 비과세되는 세제혜택을 받는다.

- <통화안정기금: CV\_FUNDS>은 통상 중앙은행의 통화안정계정에 계상·운영된다.
- 방송법 제36조 <방송발전기금: CV\_FUNDS>의 설치에 관한 규정에 따라, 방송위원회가 방송진흥 사업 및 문화·예술 진흥사업을 위하여 설치 운용하는 자금을 말한다.
- 또 해외차관, 정부보유 공기업주식, <공공자금관리기금: CV\_FUNDS>, 한국은행 등에 의해 우회적으로 투입된 자금도 넓은 의미의 <공적자금: CV\_FUNDS>으로 본다.

#### 7.8.4. 주의 사항

- (상품명) 금융 상품명으로 쓰였을 경우, AFW\_SERVICE\_PRODUCTS로 태깅한다.
  - 높은 수익률을 원하신다면 <필승코리아증권투자신탁: AFW\_SERVICE\_PRODUCTS>을 추천드립니다.

### 7.9. CV\_ART

#### 7.9.1. 정의 및 설명

- 예술의 분류, 종류
  - 문예, 음악, 미술, 춤, 영화, 드라마의 종류

#### 7.9.2. 분석 대상

- 예술의 분류
  - 문예 종류: 소설, 시, 신화
  - 소설 종류: 고전소설, 추리소설, 우화, 판타지소설, SF소설, 일본소설
  - 음악 종류: 현악 4중주, 아악, 군악, 범패, 가곡, 힙합, 팝송, 현대 음악, 소나타, 캐럴, 가요, 프랑스음악
  - 미술 종류: 동양화, 서양화, 한국화, 수채화, 풍경화, 조각, 소묘, 크로키
  - 춤 종류: 발레, 재즈댄스, 검무, 사교댄스
  - 영화 종류: 공포영화, 코미디영화, 과학영화
  - 드라마 종류: 영국드라마, 미국드라마

#### 7.9.3. 예문

- 예술의 분류, 종류
  - 한국어로 보통 <시: CV\_ART>라고 할 때에는 그 형식적 측면을 주로 가리켜 문학의 한 장르로서의 <시: CV\_ART>작품(poem)을 말하는 경우가 있다.
  - 풋풋한 목소리의 <캐럴: CV\_ART> 음반도 있다.
  - 18세기에 <소나타: CV\_ART>는 특정한 형식을 갖추게 되면서 몇 개의 악장으로 이루어진 기악의 독주곡이 되었다.

- 자, 그러면 <코미디: CV\_ART> 부문 후보를 만나볼 텐데요.

#### 7.9.4. 주의 사항

- 문예, 소설, 음악, 미술, 춤, 영화, 드라마와 같은 상위 부류는 태깅하지 않는다.
- '노래, 춤, 음악'은 보통명사로 태깅 대상이 아니다. 세부 유형인 '뮤지컬, 발레, 재즈댄스, 검무, 사교댄스' 등을 CV\_ART로 태깅한다.
  - 인기 절정의 가요보다는 음악인의 냄새가 담긴 그런 음악을 사랑하는 음악창고에 제가 오늘 초대이 되었습니다. → 인기 절정의 <가요: CV\_ART>보다는 음악인의 냄새가 담긴 그런 음악을 사랑하는 음악창고에 제가 오늘 초대이 되었습니다.
    - '음악'은 보통명사로 태깅하지 않는다.
- 예술의 장르가 아닌 예술 활동의 '결과물(작품)'은 CV\_ART로 태깅하지 않는다. 맥락상 '결과물'의 의미로 쓰인 보통명사는 태깅하지 않으며, 결과물의 이름은 AFA로 태깅한다.
  - 사진 700장을 발견했다. 김영하 작가의 신작 소설
    - 맥락상 '사진, 소설'은 예술 활동의 결과물이기 때문에 태깅하지 않는다.
  - 그가 <시: CV\_ART> '만용에게'에서 냈두리했다.
    - '만용에게'는 작품명이므로 AFA\_DOCUMENT로 태깅한다.
- (국가+예술) '국가+예술 분야'의 경우 최장 단위 태깅 원칙에 따라 '국가+예술 분야' 전체를 CV\_ART로 태깅한다.
  - 미국드라마, 프랑스음악, 일본소설 → <미국드라마: CV\_ART>, <프랑스음악: CV\_ART>, <일본소설: CV\_ART>
  - 미드 → <미드: CV\_ART>
- (K-팝, 케이팝) '케이팝'은 우리말샘에 '전 세계적인 인기를 받는 현대 한국의 대중가요를 통틀어 이르는 말'로 등재되어 있다. 팝송과 동일하게 취급하여 CV\_ART로 태깅한다.

### 7.10. CV\_SPORTS

#### 7.10.1. 정의 및 설명

- 스포츠/레포츠/레저 명칭 및 종목

#### 7.10.2. 분석 대상

- 스포츠/레포츠/레저 명칭 및 종목
  - 스케이트, 스키, 하이킹, 행글라이딩, 인라인스케이팅, 축구, 마라톤 등

#### 7.10.3. 예문

- 일상생활에서 교외로 벗어나는 산책 나들이에서부터 가벼운 등산 등 야외활동을 널리 <하

이킹: CV\_SPORTS>이라 한다.

- <행글라이딩: CV\_SPORTS>은 공기 속을 움직이는 스포츠이므로 기상에 민감한 편인데 기상조건으로는 바람이 특히 중요하다.
- 미국의 <롤러스케이팅: CV\_SPORTS>, <스피드스케이팅: CV\_SPORTS> 경기가 쿼드(quad)와 인라인(inline)으로 나뉘는 등 1990년대 들어 세계적으로 <인라인스케이팅: CV\_SPORTS> 붐이 일어났다.
- <축구: CV\_SPORTS>는 상대팀과 대결하여 득점을 겨루는 것이 목표이기 때문에 그 목표를 달성하기 위해서는 기본기술이 절대적으로 중요하다.
- <마라톤: CV\_SPORTS> 기록은 코스마다 조건이 약간 다르기 때문에 세계기록은 공인되지 않으며, '세계 최고'라는 말이 쓰인다.

#### 7.10.4. 주의 사항

- (전통놀이, 민속놀이의 종류) 전통놀이, 민속놀이의 종류는 CV\_SPORTS로 태깅한다.
  - 널뛰기, 연날리기, 칠교놀이, 낫다리밟기, 사물놀이, 윷놀이, 자치기, 제기차기 등
- (일반 놀이의 종류) 전통놀이, 민속놀이의 종류가 아닌 달리기, 줄넘기, 공놀이, 공기놀이 등은 태깅하지 않는다.
- (게임의 종류) 일상에서 게임의 형태로 많이 행해지기는 하지만 전문 종목이며 정식 스포츠에 해당되는 경우, CV\_SPORTS로 태깅한다.
  - 바둑, 당구, 포켓볼 등
- (게임의 종류) 게임으로만 행해지며 전문 종목, 스포츠로 볼 수 없는 경우는 태깅하지 않는다.
  - 빙고 게임 등

### 7.11. CV\_SPORTS\_POSITION

#### 7.11.1. 정의 및 설명

- 스포츠 포지션

#### 7.11.2. 분석 대상

- 스포츠 포지션
  - 농구 포지션: 가드, 포워드, 센터, 포인트 가드, 슈팅 가드, 스몰 포워드, 파워 포워드
  - 야구 포지션: 외야수, 좌익수, 중견수, 우익수, 내야수, 3루수, 유격수, 2루수, 1루수, 투수, 포수, 타자, 4번 타자
  - 축구 포지션: 수비수, 미드필더, 공격수, 골키퍼
  - 배구 포지션: 리베로, 세터, 레프트, 라이트, 센터

- 그 밖의 스포츠 포지션: 센터 포워드, 레프트 윙, 라이트 윙, 레프트 백, 라이트 백, 센터 백, 레프트 디펜스, 라이트 디펜스 등

### 7.11.3. 예문

- <슈팅 가드: CV\_SPORTS\_POSITION>는 과거에도 그렇고 지금도 팀의 에이스를 담당하는 선수들이 많이 있습니다.
- 보통 KBO 최고의 <타자: CV\_SPORTS\_POSITION>는 이대호, 한국 야구 최고의 <타자: CV\_SPORTS\_POSITION>는 추신수가 거론된다.
- 아르헨티나의 베테랑 <미드필더: CV\_SPORTS\_POSITION> 하비에르 마스체라노가 은퇴를 선언했다.
- <리베로: CV\_SPORTS\_POSITION>는 축구와 배구에서 선수 포지션을 나타내는 용어로 <수비수: CV\_SPORTS\_POSITION>의 일종이다.

### 7.11.4. 주의 사항

- (4번 타자) 사전에 '야구에서, 네 번째 타자를 이르는 말. 팀에서 타격 실력이 월등한 선수로, 팀의 중심 타자를 가리킨다'고 나와 있다. '4번 타자', '사번 타자'의 경우, CV\_SPORTS\_POSITION으로 태깅한다.

## 7.12. CV\_SPORTS\_INST

### 7.12.1. 정의 및 설명

- 스포츠/레포츠/레저의 용품/도구 명칭

### 7.12.2. 분석 대상

- 스포츠/레포츠/레저의 용품/도구 명칭
  - 스케이트, 카빙스키, 행글라이더 등

### 7.12.3. 예문

- 레저용으로 흔히 볼 수 있는 <롤러블레이드: CV\_SPORTS\_INST>(<roller blade: CV\_SPORTS\_INST>)의 공식명칭인 <인라인스케이트: CV\_SPORTS\_INST>(<inline skate: CV\_SPORTS\_INST>)는 한 줄로 된 프레임에 3~5개의 바퀴가 일렬로 고정되어 있다.

### 7.12.4. 주의 사항

- (운동 의복) 수영복, 등산복 등의 운동 의복은 CV\_SPORTS\_INST로 태깅하지 않는다. CV\_CLOTHING으로 태깅한다.

- (상품명) 스포츠/레포츠/레저의 용품/도구 명칭이 상품명으로 쓰였을 경우, AFW\_OTHER\_PRODUCTS로 태깅한다. 보통, AFW\_OTHER\_PRODUCTS의 경우는 스포츠/레저 용품의 상품명으로 되어 있어 구분이 어렵지 않다.
  - 기왕이면 <삼천리: AFW\_OTHER\_PRODUCTS> 사 보려구 하는데 역시 <자이언트: AFW\_OTHER\_PRODUCTS>가 가성비 좋구나.

## 7.13. CV\_PRIZE

### 7.13.1. 정의 및 설명

- 상과 훈장

### 7.13.2. 분석 대상

- 상과 훈장
  - 노벨경제학상, 노벨문학상, 노벨물리학상, 노벨생리의학상, 노벨평화상, 노벨화학상 등
  - 대통령상, 경제정의기업상, 자기상표제품 수출상, 대중상

### 7.13.3. 예문

- <대중상: CV\_PRIZE>은 한국영화계의 대표적 영화제로 존속하면서 우수영화의 발굴 및 포상을 통한 진흥효과를 거둔 것은 긍정적이지만

### 7.13.4. 주의 사항

- (대상 등 상장명) '대상', '최우수상', '최고상', '우수상' 등의 보통명사도 CV\_PRIZE로 태깅한다. 그러나 '우승, 준우승'은 CV\_PRIZE로 태깅하지 않는다.
  - 대상, 최우수상, 최고상, 우수상, 인기상
- 상과 훈장의 명칭은 DATE, QUANTITY와 함께 쓰이는 경우가 많다. 이 경우 DATE와 QUANTITY는 분리하여 태깅한다.
  - 2000년 제34회 납세자의 날 대통령상 → <2000년: DT\_YEAR> <제34회: QT\_ORDER> <납세자의 날: DT\_DAY> <대통령상: CV\_PRIZE>
  - 제10회 경제정의기업상 → <제10회: QT\_ORDER> <경제정의기업상: CV\_PRIZE>

## 7.14. CV\_RELATION

### 7.14.1. 정의 및 설명

- 가족/친족의 구성원



- 가족/친족 관계를 지칭하는 명사
- 유족, 유가족

#### 7.14.2. 분석 대상

- 가족/친족의 구성원
  - 아버지, 어머니, 할아버지, 할머니, 아들, 딸, 언니, 오빠, 형, 동생, 신랑, 남편, 아내, 마누라, 시어머니, 사위, 엄빠
- 가족/친족 관계를 지칭하는 명사
  - 부부, 부모, 형제, 자매, 남매, 부자, 모녀, 친척
  - 유족, 유가족

#### 7.14.3. 예문

- 가족/친족의 구성원
  - 은경은 <어머니: CV\_RELATION>의 간청으로 실어증에 걸린 <아버지: CV\_RELATION>를 만나고 <아버지: CV\_RELATION>에게 연민의 정을 느낀다.
  - 즉 <딸: CV\_RELATION>만 있는 집에서 <데릴사위: CV\_RELATION>로 삼아 같이 살면서 그의 노동력(생산력)으로 생활하는
- 가족/친족 관계를 지칭하는 명사
  - 결혼 날짜를 정하기 위해 양가 <부모: CV\_RELATION>들이 만남을 가졌다.
  - 비봉촌 창윤이네 집에 남게 된 것은 창윤이 <부부: CV\_RELATION>와 정수 <남매: CV\_RELATION> 네 식구밖에 없었다.
  - 준의 <유족: CV\_RELATION>은 오후까지 오열하며 국립묘지를 떠나지 않았다.

#### 7.14.4. 주의 사항

- (가족, 아이) '가족, 아이'의 경우, 가족의 구성원이나 가족 관계를 나타낼 때에는 CV\_RELATION으로 태깅한다.
  - 너의 가족에게도 통지가 갔을 것이었다. → 너의 <가족: CV\_RELATION>에게도 통지가 갔을 것이었다.
    - 전시회에 가족 단위의 관람객들이 줄을 잇고 있다. → 가족 관계를 나타내지 않으므로 CV\_RELATION으로 태깅하지 않는다.
  - 4살 유아가 있는 우리집은 다양한 동물들에게 밥 줄 수 있어 아이가 참 좋아했다. → 4살 유아가 있는 우리집은 다양한 동물들에게 밥 줄 수 있어 <아이: CV\_RELATION>가 참 좋아했다.
    - 먹는 것을 가지고 어른이 아이를 울리면 어떻게 해요? → 가족 관계를 나타내지 않으므로 CV\_RELATION으로 태깅하지 않는다.
- (예비신부, 예비신랑, 예랑) 예비신랑, 예비신부의 경우 가족 구성원의 의미를 나타낸다고

볼 수 있으므로 CV\_RELATION으로 태깅한다.

- 내 예비신랑을 소개해 볼까 해요. → 내 <예비신랑: CV\_RELATION>을 소개해 볼까 해요.
- 예비신랑님, 들어오십니다. → <예비신랑: CV\_RELATION>님, 들어오십니다.
- '사촌동생, 사촌오빠'의 경우 우리말샘에 구 단위로 등재되어 있으나 의미를 고려하여 전체를 CV\_RELATION으로 태깅한다. '사촌형제'는 사전에 등재되어 있지 않으나 같은 기준을 적용하여 전체를 CV\_RELATION으로 태깅한다.
  - 사촌동생 → <사촌 동생: CV\_RELATION>, 사촌오빠 → <사촌 오빠: CV\_RELATION>
  - 사촌형제 → <사촌 형제: CV\_RELATION>
- '음악의 아버지'와 같이 가족/친족의 대상이 아닌 경우는 CV\_RELATION으로 태깅하지 않는다.
  - 바흐가 음악의 아버지라고 불리는 이유는 바로크 시대 이전의 음악이 오늘날의 서양 음악으로 발전할 수 있도록 음악의 기초를 닦았기 때문이지. → 여기서의 '아버지'는 태깅 대상이 아니다.

## 7.15. CV\_OCCUPATION

### 7.15.1. 정의 및 설명

- 직업

### 7.15.2. 분석 대상

- 직업
  - 공공 기관 종사자
    - 공무원
  - 전문가 및 관련 종사자
    - 연구원, 컨설턴트, 프로그래머, 의사, 한의사, 간호사, 수의사, 약사, 영양사, 위생사, 목사, 신부, 승려, 교수, 시간강사, 교사, 조교, 판사, 검사, 변호사, 변리사, 회계사, 세무사, 관세사, 작가, 기자, 평론가, 번역가, 통역가, 감독, 배우, 성우, 개그맨, 아나운서, 쇼핑호스트, 리포터, 화가, 사진가, 성악가, 가수, 무용가, 디자이너, 감독, 코치, 선수, 매니저, 파일럿, 조종사
  - 사무 종사자
    - 사무원, 비서, 데스크 안내원, 속기사
  - 서비스 종사자
    - 경찰관, 소방관, 교도관, 경호원, 미용사, 분장사, 메이크업 아티스트, 웨딩플래너, 장례 지도사, 점술가, 카지노 딜러, 캐디, 조리사, 바텐더, 웨이터, 가이드, 관광가이드

- 판매 종사자
  - 영업원, 보험 설계사, 텔레마케터, 방문 판매원
- 농림, 어업 숙련 종사자
  - 재배원, 조경원, 사육사, 감별사, 어부, 해녀
  - 기능원 및 관련 기능 종사자
    - 제과원, 도축원, 재단사, 재봉사, 전기공, 도배공, 배관공
  - 단순노무 종사자
    - 배달원, 택배기사, 환경미화원, 경비원, 가사 도우미
  - 군인
  - 기타
    - 학생, 탐정, 큐레이터, 나레이터모델, 외환딜러, 브로커, 스파이, 오피상, 마도로스, 장돌뱅이, 기공, 기수(騎手)

### 7.15.3. 예문

- 이 <공인중개사: CV\_OCCUPATION>는 정부가 부동산중개업자의 공신력(公信用)을 높이고 공정한 부동산 거래질서를 확립한다는 입법취지로,
- 단체급식의 관리책임자는 <영양사: CV\_OCCUPATION>가 되어야 하며,
- 부대에 배치된 <군인: CV\_OCCUPATION>은 참모요원으로서 부대장병의 건강관리와 위생에 관한 직무를 수행하며 부대장을 보좌한다.
- <아나운서: CV\_OCCUPATION>는 방송국명의 고지, 방송순서의 소개, 뉴스 방송, 스포츠나 식전의 실황중계, 대담의 사회, 낭독 등 다방면에 걸쳐 방송순서를 구성하는 중요한 요소가 된다.
- 일찍이 제1차 세계대전 당시 이탈리아 전선에 군속으로 종군하여 부상을 바 있는 E.M.헤밍웨이는 <종군작가: CV\_OCCUPATION>의 대표적인 예이다.

### 7.15.4. 주의 사항

- 적정술어 '-으로 일하다', '-을 해서 먹고 살다', '-을 해서 생계를 유지하다'와 어울릴 경우, CV\_OCCUPATION으로 태깅한다.
- (학생) '학생'은 위의 적정술어와 어울리지 않지만 '직업란'에 '학생'을 쓰는 경우가 있으므로 예외적으로 CV\_OCCUPATION으로 태깅한다.
  - 이외에 '대학생, 고등학생, 중학생, 초등학생' 등은 태깅 대상이 아니다.
- (선생님) '교사'와 같이 직업을 나타낼 경우만, CV\_OCCUPATION으로 태깅한다.
  - 김철수 선생님 되시지요? 여기 한라은행 춘천점입니다. → 남을 대접하여 이르는 말로 사용한 '선생님'의 경우는 태깅 대상이 아니다.
  - 우리 동네 의사 선생님 참 친절하다. → '직업명+선생님'에서 '선생님'은 '교사'가 아니므로 태깅 대상이 아니다.

- (운동 종목+선수) 우리말샘 사전의 분석에 따라 최장 단위로 태깅한다.
  - 축구 선수 → <축구 선수: CV\_OCCUPATION>
    - 우리말샘 검색 결과 '축구 선수'의 경우, '축구^선수'로 등재되어 있다. '^'으로 표시된 경우, 전체를 하나로 태깅한다.
- 유투버는 최근 초등학생 장래희망직업 순위에 오르기도 했다. 새로 생겨난 말 중에 직업으로 인정이 되면서 우리말샘에 등재되어 있으면 CV\_OCCUPATION으로 태깅한다.
  - 유투버, 블로거, 파워블로거 → CV\_OCCUPATION으로 태깅한다.
    - '뷰티 블로거'는 사전에 등재되어 있지 않다. 이 경우에는 '블로거'만 태깅 대상이 된다.
      - 뷰티 블로거 → 뷰티 <블로거: CV\_OCCUPATION>
  - '웨딩플래너, 카지노 딜러, 가사도우미' 등 몇몇 직업 명칭은 우리말샘에 구 단위로 등재되어 있으나 의미를 고려하여 전체를 CV\_OCCUPATION으로 태깅한다.
    - 파워 블로거 → <파워 블로거: CV\_OCCUPATION>

## 7.16. CV\_POSITION

### 7.16.1. 정의 및 설명

- 공무원의 직위/직책/직급  
(행정부, 지자체, 입법부, 사법부, 검찰, 경찰, 군대)
- 군인의 직위/직책/직급
- 직장에서의 직위/직책/직급
- 교수의 직위/직책/직급
- 학교에서의 직위/직책/직급
- 고대 국가의 직위/직책/직급

### 7.16.2. 분석 대상

- 공무원의 직위/직책/직급
  - (행정부) 대통령, 국무총리, 장관, 대사, 국장, 과장, 계장, 주무관, 주사보, 서기, 서기보
  - (지자체) 시장, 도지사, 광역시장, 부시장, 군수, 부군수, 구청장
  - (입법부) 국회의장, 국회부의장, 국회의원
  - (사법부) 대법원장, 헌법재판소장, 고등법원장, 사법연수원장, 부장판사
  - (검찰) 검찰총장, 고검장, 지검장, 차장검사, 부장검사
  - (경찰) 경찰청장, 경찰청차장, 경무관, 총경, 경찰서장, 경감, 경위, 경사, 경장, 순경
  - (군대) 대장, 중장, 소장, 대령, 중령, 소령, 대위, 중위, 소위, 장교, 군의관, 상사, 중사, 하사, 부사관

- 군대에서의 직위/직책/직급
  - 병장, 상병, 일병, 이병
- 직장에서의 직위/직책/직급
  - 대표이사, 회장, 사장, 부장, 차장, 과장, 계장, 대리, 사원
  - 본부장, 팀장, 파트장
- 교수의 직위/직책/직급
  - 정교수, 부교수, 조교수
- 학교에서의 직위/직책/직급
  - 총장, 학장, 학과장, 교장, 교감
- 고대 국가의 직위/직책/직급
  - 황제, 장군, 왕, 영의정, 1품, 대료, 귀족, 군주

### 7.16.3. 예문

- 공무원의 직위/직책/직급
  - 지자체 또는 <주무관: CV\_POSITION>들마다 다른 업무처리부터 시급히 해결해야 합니다.
  - 김 <부시장: CV\_POSITION>은 빠른 판단과 꼼꼼한 성격으로 빠르고 정확하게 업무를 추진할 뿐만 아니라 평소 소통과 겸손을 중요시 여겨 직원들과도 가깝게 지낸다.
  - <국회의원: CV\_POSITION>은 국민을 대표하여 법률을 제정하고 국정을 심의한다.
  - 현직 <부장판사: CV\_POSITION>가 서울 강남의 한 술집에서 술값 시비를 벌이다 종업원과 경찰관을 폭행해 입건됐다.
  - <경위: CV\_POSITION>는 경찰 공무원 계급의 하나로, <경감: CV\_POSITION>의 아래, <경사: CV\_POSITION>의 위이다.
  - <하사: CV\_POSITION>는 <부사관: CV\_POSITION> 계급에서 가장 낮은 계급이다.
- 직장에서의 직위/직책/직급
  - <사장: CV\_POSITION>·<부사장: CV\_POSITION> 등은 최고관리자, <부장: CV\_POSITION>·<과장: CV\_POSITION> 등은 중간관리자, <계장: CV\_POSITION>은 하급관리자이다.
  - 현 <대표이사: CV\_POSITION>가 임기가 만료되어 퇴임 등기를 신청하면서, 새로운 사람을 <대표이사: CV\_POSITION>로 취임 등기하려고 합니다.
- 교수의 직위/직책/직급
  - 대학에 임용되는 대부분의 교수들은 <조교수: CV\_POSITION>로 시작한다.
- 학교에서의 직위/직책/직급
  - 14일 <학과장: CV\_POSITION> 회의를 열고 미등록 학생 처리 문제를 논의한 끝에 학교에서 등록금을 대출해 주는 등의 방법으로 17일까지 구제 활동을 계속하기로 했다.
- 고대 국가의 직위/직책/직급

- 15세에 음보(蔭補)로 <교위: CV\_POSITION>(校尉)가 되고, 뒤에 <장군: CV\_POSITION>에 이르렀다.
- 고대의 <왕: CV\_POSITION>은 입법·사법·행정 등의 모든 국가권력을 장악하고 군사상 통수권을 가지고 있었으며 종교적으로 신격화되었다.

#### 7.16.4. 주의 사항

- (사람 이름+직위) 사람 이름+직위의 경우 각각 분리하여 태깅한다.
- 류샤오밍 대사 → <류샤오밍: PS\_NAME> <대사: CV\_POSITON>
- (지역명+직위) 시장, 도지사, 광역시장, 부시장, 군수, 부군수, 구청장 등의 지차체 공무원 직위/직책/직급의 경우, 지역명이 함께 쓰이면 우리말샘 사전의 분석에 따라 최장 단위로 태깅한다.
  - 광주 시장 → <광주 시장: CV\_POSITON>, 사하구청장 → <사하구청장: CV\_POSITON>
- (기관+직위) '환경부 장관', '남구청 부구청장', '선거대책위원장' 등과 같이 기관과 직위가 함께 쓰인 경우, 사전에는 등재되어 있지 않으나 기관과 직위를 나누기 어려운 경우가 있으므로 전체를 CV\_POSITION으로 태깅한다.
  - 환경부 장관 → <환경부 장관: CV\_POSITON>
  - 남구청 부구청장 → <남구청 부구청장: CV\_POSITON>
  - 선거대책위원장 → <선거대책위원장: CV\_POSITON>
- (부서+직위) '환경정책과장' 등과 같이 부서와 직위가 함께 쓰인 경우, 사전에는 등재되어 있지 않으나 부서와 직위를 나누기 어려운 경우가 있으므로 전체를 CV\_POSITION으로 태깅한다.
  - 환경정책과장 → <환경정책과장: CV\_POSITON>
- 현 사장, '전 국무총리'와 같이 직위에 '전, 현'이 포함되어 나올 경우 '전, 현'은 태깅하지 않는다.
  - 현 사장 → 현 <사장: CV\_POSITON>
- 왕의 명칭의 경우, 묘호와 직위 및 직책을 분리하여 태깅한다.
- 조선 3대 태종왕 이방원의 묘 헌릉에 다녀왔어요. → 조선 3대 <태종: PS\_NAME><왕: CV\_POSITON> 이방원의 묘 헌릉에 다녀왔어요.

### 7.17. CV\_FOOD

#### 7.17.1. 정의 및 설명

- 음식의 유형
- 요리명
- 식품명

- 식재료

### 7.17.2. 분석 대상

- 음식의 유형
  - 국, 탕국, 밥, 반찬, 국수, 나물, 무침, 향신료, 안주, 술안주
- 요리명
  - 일반 요리명: 쇠고기 덮밥, 어리굴젓, 명란젓, 해물겨자채, 김치, 나박김치, 봄배추김치, 짬지, 오이소박이, 열무김치, 순살치킨, 양념치킨
  - 식당, 음식점 등의 메뉴명: 붉은대게다리살죽, 어니언치킨, 파닭, 1955버거, 해피밀
- 식품명
  - 진짬뽕, 신라면, 초코파이, 새우깡, 연두, 멸치디포리다시팩, 일판란, 냉동차돌박이바로구이, 새콤달콤꼬마유부초밥
- 식재료
  - 국물의 명칭: 현미, 좁쌀, 율무, 백미
  - 음식 재료: 소고기, 돼지고기, 닭고기, 버섯, 열무, 파
  - 양념: 소금, 설탕, 간장, 겨자즙

### 7.17.3. 예문

- 음식의 유형
  - 한국에서도 아주 오래 전부터 <국수: CV\_FOOD>를 만들어 먹었으나, 밀의 생산이 많지 않았기 때문에 상용음식이 되지는 못하였다.
  - <나물: CV\_FOOD>은 그 중 숙채류의 대표적인 요리라 할 수 있다.
  - 요리명
    - 특히 계절에 따라 여러 가지 <김치: CV\_FOOD>를 담가 먹었는데, 봄철에는 <나박김치: CV\_FOOD> · <봄배추김치: CV\_FOOD> · <짬지: CV\_FOOD>를 주로 담가 먹었고, 여름에는 <오이소박이: CV\_FOOD> · <열무김치: CV\_FOOD>가 주류를 이루었다.
  - 맥도날드에서 '<1955버거: CV\_FOOD>'를 새롭게 출시했다.
- 식품명
  - 오리온 '<초코파이: CV\_FOOD>'가 외국인들이 뽑은 맛있는 한국 과자 1위에 선정되었다.
  - 영화의 인기에 힘입어 '<짜파게티: CV\_FOOD>'와 '<너구리: CV\_FOOD>'를 한 묶음으로 팔고 있다.
- 식재료
  - <현미: CV\_FOOD>는 정백으로 인한 영양분의 손실이 없으므로 <백미: CV\_FOOD>에 비하여 지방, 단백질, 비타민 B1·B2 가 풍부하다.
  - <해물겨자채: CV\_FOOD>는 재료도 좋아야 하지만 <겨자즙: CV\_FOOD>의 간을 잘 맞추

고 넣는 양이 적당해야 맛있다.

#### 7.17.4. 주의 사항

- 음식 재료명으로 쓰였다고 판단되는 경우 AM 또는 PT에 속하는 개체라 할지라도 CV\_FOOD로 태깅한다.
  - 저녁 반찬으로 <시금치: CV\_FOOD>를 먹었다.
  - <시금치: PT\_GRASS>는 높이 30~60cm로 뿌리는 굵고 붉으며, 잎은 어긋나고 세모진 달걀 모양이다.
  - <젤라토: CV\_FOOD>부터 <장미아이스크림: CV\_FOOD>까지 종류도 다양하고 맛도 좋은 아이스크림 맛집으로 초대합니다.
- (밥) '음식'의 의미인 '밥'과 '끼니, 식사'의 의미인 '밥' 모두 CV\_FOOD로 태깅한다.
- (곡물+밥) 쌀밥, 보리밥 등의 경우, 우리말샘에 '쌀-밥', '보리-밥'으로 등재되어 있다. 이 경우, 전체를 CV\_FOOD로 태깅한다.
- (삶은 계란, 삶은 달걀) 관형사형 어미와 함께 쓰인 '삶은 계란, 삶은 달걀'의 경우는 우리말샘 사전에 구 단위로 등재되어 있다. 따라서 관형사형 어미를 제외한 '계란, 달걀'만을 CV\_FOOD로 태깅한다.
  - 삶은 계란 → 삶은 <계란: CV\_FOOD>
- (가공 식품) 생선이나 과일을 건조시킨 건어물, 건과일 등은 식용을 목적으로 식재료를 가공한 것이므로 CV\_FOOD로 태깅한다.
  - 북어, 건태, 건명태, 건포도, 꽃감
- (식당, 음식점 메뉴) 식당, 음식점 메뉴 중 요리명에 해당하는 것은 모두 CV\_FOOD로 태깅한다. 단, 식당의 음식 메뉴로 세트의 개념인 '모듬한판'과 같은 경우, 요리명으로 볼 수 없기에 태깅하지 않는다. '한판'만 QT로 태깅한다.
  - 푸라닭 치킨 → <푸라닭 치킨: CV\_FOOD>
  - 시간의 정성 진카스테라 → <시간의 정성 진카스테라: CV\_FOOD>
- (식품명) 마트, 인터넷마켓 등에서 판매되는 모든 '식품'(상품명)은 CV\_FOOD로 태깅한다.
  - 밤에 배가 고파서 라면을 먹었다. → 밤에 배가 고파서 <라면: CV\_FOOD>을 먹었다.
  - 요즘 참깨라면이 맛있어서 자주 먹어. → 요즘 <참깨라면: CV\_FOOD>이 맛있어서 자주 먹어.

### 7.18. CV\_DRINK

#### 7.18.1. 정의 및 설명

- 음료의 하위 부류
- 주류의 하위 부류



- 상품명

### 7.18.2. 분석 대상

- 음료의 하위 부류
  - 주스, 우유, 요구르트, 홍차, 레모네이드, 콜라, 사이다, 물
- 주류의 하위 부류
  - 위스키, 칵테일, 샴페인, 소주, 맥주, 양주, 술
- 상품명
  - 코카콜라, 칠성사이다, 카스, 테라
- 앳!메리카노, 뽕's 라떼

### 7.18.3. 예문

- 음료의 하위 부류
  - <주스: CV\_DRINK>의 원료는 오렌지·레몬·귤·사과·포도·파인애플·토마토·당근 등이다.
  - 동양 사람들은 차의 빛깔이 붉다고 하여 <홍차: CV\_DRINK>라고 하지만 영어로는 차잎이 검다고 하여 <블랙 티: CV\_DRINK>(<黑茶: CV\_DRINK>)라고 한다.
  - <레모네이드: CV\_DRINK>는 레몬과 <에이드: CV\_DRINK>를 합친 말이며, <에이드: CV\_DRINK>란 천연과즙에 설탕과 물을 섞었다는 뜻이다.
- 주류의 하위 부류
  - 이와 같이 <칵테일: CV\_DRINK>은 미국에서 처음 만들기 시작했다고 하나 혼성음료를 만드는 습관은 반드시 미국에서 시작된 것은 아니고,
  - <샴페인: CV\_DRINK>은 마개가 빠질 때 나는 평하는 소리와 함께 이는 거품이 특징인 <술: CV\_DRINK>이다.
  - 정통 <폭탄주: CV\_DRINK>는 <맥주: CV\_DRINK>에 <위스키: CV\_DRINK>를 혼합한 <칵테일: CV\_DRINK>이다.
- 상품명
  - <코카콜라: CV\_DRINK>부터 얼음컵에 팔팔팔 따라 봅니다.
  - 원래 술 마시는 거 안 좋아하는데 <테라: CV\_DRINK> 왜 이리 맛있나요?
  - <앳!메리카노: CV\_DRINK> 다 마셔갈 때 찰칵했네요.

### 7.18.4. 주의 사항

- (음료, 음료수) 주스, 우유, 콜라 등의 상위 개념인 '음료, 음료수'는 태깅 대상이 아니다.
- (물) 마시는 용도의 '물'은 음료의 하위 부류이다. 마시는 용도로 '물'이 사용된 경우, CV\_DRINK로 태깅한다.
  - 저는 <커피: CV\_DRINK> 괜찮습니다. <물: CV\_DRINK> 주세요.

- 수영을 배운 첫날이라서 물을 많이 먹었어. → 여기에서 '물'은 태깅 대상이 아니다.
- (상품명) 마트, 인터넷마켓 등에서 판매하는 모든 음료명, 주류명은 CV\_DRINK로 태깅한다.  
또한 '앗!메리카노, 뽕's 라떼'와 같은 카페 등의 모든 음료 메뉴도 CV\_DRINK로 태깅한다.
- (~ 주스) '에이드'와 '레모네이드'를 모두 CV\_DRINK로 태깅하는 것과 마찬가지로, 주스의 다양한 종류인 오렌지주스, 토마토주스 등도 CV\_DRINK로 태깅한다.
- 오렌지 주스 → <오렌지 주스: CV\_DRINK>

## 7.19. CV\_FOOD\_STYLE

### 7.19.1. 정의 및 설명

- 음식의 종류

### 7.19.2. 분석 대상

- 음식의 종류
  - 한식, 일식, 양식, 중식
  - 보양식, 패스트푸드, 육식, 채식, 한방 식품
  - 한국음식, 중국음식, 일본음식, 미국음식

### 7.19.3. 예문

- <양식: CV\_FOOD\_STYLE>보단 <한식: CV\_FOOD\_STYLE>이 내 입맛에 맞는다.
- 요즘처럼 불볕더위에는 냉면이나 빙수 등 차가운 음식만을 찾기가 쉽지만 땀을 많이 흘려 소진된 기력을 회복하는 데는 삼계탕 등 뜨거운 <보양식: CV\_FOOD\_STYLE>이나 <한방 식품: CV\_FOOD\_STYLE>의 섭취가 필수적이다.
- 홍합은 <고단백 저지방 식품: CV\_FOOD\_STYLE>이면서
- 콜로이드와 단백질이 풍부하여 <여름보양식: CV\_FOOD\_STYLE>으로 유명한 삭스핀을 먹었다.
- <중국음식: CV\_FOOD\_STYLE>의 맛은 전통적으로 오미 즉, 짠맛, 매운맛, 신맛, 단맛, 쓴맛으로 분류된다.

### 7.19.4. 주의 사항

- '절식'은 음식 종류가 아니므로 CV\_FOOD\_STYLE로 태깅하지 않는다.
- '야식, 간식, 디저트, 스낵, 국물'은 CV\_FOOD\_STYLE로 태깅하지 않는다. 개체명으로 보기 어렵다.

## 7.20. CV\_CLOTHING

### 7.20.1. 정의 및 설명

- 섬유의 명칭
- 의복의 명칭
- 의복의 유형
- 모자, 가방, 양말, 신발의 하위 유형

### 7.20.2. 분석 대상

- 섬유의 명칭
  - 모시, 레이온, 양모, 면, 마, 견, 모, 메리야스, 기모, 나일론
- 의복의 명칭
  - 상의 명칭: 와이셔츠, 티셔츠
  - 하의 명칭: 카르고 팬츠, 배기팬츠, 청바지
- 의복의 유형
  - 교복, 세라복, 등산복, 수영복
- 모자, 가방, 양말, 신발의 하위 유형
  - 베레모, 비니, 페즈
  - 숄더백, 크로스백, 백팩, 토트백, 버킷백
  - 루즈 삭스, 페이크삭스, 무지양말, 스타킹, 버선
  - 등산화, 축구화, 하이힐, 장화, 고무신, 슬리퍼, 샌들

### 7.20.3. 예문

- 섬유의 명칭
  - <레이온: CV\_CLOTHING>의 큰 결점은 내수성(耐水性)이 크게 떨어진다는 것과, 탄성이 적어 의복재료로는 구김살이 생기기 쉽다는 점이다.
  - <양모: CV\_CLOTHING>는 오스트레일리아·뉴질랜드·아르헨티나·미국·러시아 등지에서 많이 생산된다.
- 의복의 명칭
  - 특히 국방색 <카르고 팬츠: CV\_CLOTHING>는 <카무플라주 점퍼: CV\_CLOTHING>나 <셔츠: CV\_CLOTHING>를 입어 밀리터리 스타일로 연출할 수도 있다.
  - 그는 <청바지: CV\_CLOTHING>에 <티셔츠: CV\_CLOTHING>의 편한 차림을 즐겨 입는다.
- 의복의 유형
  - 방 안은 여학생이 쓰던 방인 듯 책꽂이와 함께 벽에 여학생의 <교복: CV\_CLOTHING>까지 걸려 있었다.
  - 불암사 근처에서 왓자한 인기척이 들려왔다. 해 질 녘 안으로 웅포까지 들어갈 작정인

듯 한 때의 <등산복: CV\_CLOTHING> 차림을 한 젊은이들이 걸음을 서둘러 내려오고 있었다.

- 모자, 가방, 양말, 신발의 하위 유형
  - <베레모: CV\_CLOTHING>를 납작하게 눌러쓴 얼굴이 여간 깡깡하게 보이는 게 아니었다.
  - 15인치 노트북을 갖고 나갈 일이 생긴다면 <백팩: CV\_CLOTHING>에 집어넣고 다니는 것이 몸도 마음도 편하다.
  - 장식은 금속 체인이나 벨트 장식이 많고, 80년대 영향으로 <루즈 삭스: CV\_CLOTHING>처럼 흘러내리는 주름진 스타일도 눈에 띈다.

#### 7.20.4. 주의 사항

- (대분류) '바지, 모자, 상의, 하의, 양말, 가방, 신발' 등 대분류는 태깅하지 않는다.
- (장신구) 귀걸이, 목걸이, 반지 등의 착용물은 태깅하지 않는다.
- (적정술어 사용) '-을/를 입다, 신다, 쓰다', '-을 벗다'와 어울리지 않으면 태깅하지 않는다. 옷, 신발류, 양말류, 가방류는 이러한 서술어와 어울리지만 장신구류는 전혀 어울리지 않는다.
- (섬유 명칭+바지) '섬유 명칭+바지'의 경우, 우리말샘 사전을 참고하여 다음과 같이 태깅한다. 우리말샘에서 '-바지'를 검색하면 '면-바지', '기모^바지'로 등재되어 있다. '-', '^'의 표시가 있는 경우, 전체를 CV\_CLOTHING으로 태깅한다. '마 바지', '가죽 바지'는 띄어쓰기가 된 구 구성으로 사전에 등재되어 있으나, '면-바지', '기모^바지'와 동일한 기준을 적용하여 전체를 CV\_CLOTHING으로 태깅한다.
  - 면바지 → <면바지: CV\_CLOTHING>, 기모바지 → <기모바지: CV\_CLOTHING>
  - 마 바지 → <마 바지: CV\_CLOTHING>, 가죽 바지 → <가죽 바지: CV\_CLOTHING>, 나일론 바지 → <나일론 바지: CV\_CLOTHING>

### 7.21. CV\_BUILDING\_TYPE

#### 7.21.1. 정의 및 설명

- 건축 양식 유형

#### 7.21.2. 분석 대상

- 건축 양식 유형
  - 기능주의건축, 노르만양식, 시토파건축, 에마뉘엘 양식, 그리스건축, 도리스양식
  - 유럽식 목조주택, 기와집, 널벽집, 한옥

### 7.21.3. 예문

- 현존하는 가장 오래되고 전형적인 <시토파건축: CV\_BUILDING\_TYPE>이다.
- 파타라가 <에마뉘엘 양식: CV\_BUILDING\_TYPE>의 중심지였다.
- <그리스건축: CV\_BUILDING\_TYPE>의 중심은 신전이므로 그 역사도 신전 형식의 발전에 비례한다.
- <도리스 양식: CV\_BUILDING\_TYPE>의 가장 오래된 유구(遺構)는 올림피아의 헤라신전(BC 7 세기 중엽)이며, 이 밖에 코린트의 아폴론신전(BC 5 세기 초엽), 아테네의 파르테논신전(BC 5 세기 중엽) 등이 있다.

## 8. DATE

---

DATE는 '구체적인 기간'을 나타내는 DURATION, '일(日)'을 나타내는 DAY, '주(週)'를 나타내는 WEEK, '월(月)'을 나타내는 MONTH, '해(年)'를 나타내는 YEAR, '계절'을 나타내는 SEASON, '지질시대'를 나타내는 GEOAGE, '왕조 및 그 시대'를 나타내는 DYNASTY, 그리고 불명확한 기간을 나타내는 OTHERS의 9가지 세부분류로 나뉜다.

- 주의 사항
  - DURATION과 OTHERS의 경계
    - '구체적인 기간'을 나타내는 것은 모두 DURATION으로 태깅한다. 반면, 기간을 나타내기는 하지만 구체적인 날짜를 알 수 없는 경우나 세부분류 1~8에 해당되지 않는 경우에는 OTHERS로 태깅한다.
  - 보통명사 태깅 여부
    - 오늘, 내일, 이달, 올해, 평일, 주말, 이번 주 등과 같은 보통명사 또한 태깅한다. 각각에 해당하는 세부분류는 아래 지침에서 확인할 수 있다.
  - 태깅 단위
    - DATE와 함께 '지난~, 저번~, 이번~, 오는~, 다음~, ~초, ~말, ~쯤'과 같이 날짜에 구체성을 더하는 표현이 있다면 함께 태깅한다.
  - 숫자만 쓰인 경우
    - 날짜 표현 없이 숫자 자체만으로도 문맥상 날짜를 의미하는 것이라면 DATE로 태깅한다.
  - QT\_ORDER와의 경계
    - (n일) '일'을 단위로 날의 수를 세는 경우는 QT\_COUNT, '얼마 동안'의 경우는 DT\_DURATION, 달력상의 날짜는 DT\_DAY로 태깅한다.
      - 올해 공휴일은 31일밖에 안 된다. → 올해 공휴일은 <31일 밖에: QT\_COUNT>

안 된다.

- 커플 31일 기념 프사 → <31일: DT\_DURATION> 기념 프사
  - '31일'은 31일 동안 만난 것을 기념하는 의미이므로 QT\_ORDER가 아니라 DT\_DURATION으로 태깅한다.
- 3일 5시 출발 → <3일: DT\_DAY> 5시 출발
- (n일째) '얼마 동안'의 의미로 쓰였으면 DT\_DURATION, '순차'의 의미를 나타내면 QT\_ORDER로 태깅한다.
  - 3일째 되는 날 비가 내리기 시작했다. → <3일째: QT\_ORDER> 되는 날 비가 내리기 시작했다.
  - 3일째 비가 내렸다. → <3일째: DT\_DURATION> 비가 내렸다.
- (n일 차) 日次가 '날의 차례'를 의미하므로 DT가 아닌 QT로 태깅한다. (비슷한 쓰임의 '年次'에 대한 우리말샘 해설을 보면, 그 상위어가 '차례'라고 되어 있음.)
  - 교생실습 1일 차 → 교생실습 <1일 차: QT\_ORDER>
  - 입사 8년 차 → 입사 <8년 차: QT\_ORDER>

## 8.1. DT\_DURATION

### 8.1.1. 정의 및 설명

- 문맥상으로 구체적인 날짜의 기간을 나타내는 것

### 8.1.2. 분석 대상

- ~간, ~내내, ~동안, ~가량
- ~부터 ~까지, ~에서 ~까지
- 전반기/후반기, 성수기/비수기, 환절기
- 하루, 이틀, 주말, 일주일, ~주, ~개월, ~한 해, ~년
- ~시즌
- ~째

### 8.1.3. 예문

- ~간, ~내내, ~동안, ~가량
- 부민문화사에서 발행된 산림기사 기출문제 해설은 <10년간: DT\_DURATION>의 기출문제를 모두 담아 해설을 첨부하였습니다.
- 안스리움은 <일년 내내: DT\_DURATION> 피는 꽃이다.
- 과거 <50년 동안: DT\_DURATION> 한국이 가난에서 벗어나서 경제발전을 했던 그 경험들이 있잖아요.

- ~부터 ~까지, ~에서 ~까지  
볼리비아의 에보 모랄레스 정권은 <2006년부터 2016년까지: DT\_DURATION> 지속되었다.
- 생각해 보니 <이번 달부터 다음 달까지: DT\_DURATION>는 중요한 일정으로 꼭 차서 만나기가 어려울 것 같네.
- 철학의 원조라고 하는 고대 그리스의 철학자 탈레스(<BC624~BC546: DT\_DURATION>)는 '짐승이 아니라 인간으로 태어난 것을 감사의 첫 번째로 삼았고, 두 번째는 여자가 아니라 남자로 태어난 것, 세 번째는 야만인이 아니라 그리스인으로 태어난 것'이라고 했다.
- 대릉원은 <4세기에서 6세기경: DT\_DURATION>의 신라 지배계층의 무덤이 모여 있는 곳입니다.
- 행사 기간은 <2020.1.13(수)~2020.2.15(금): DT\_DURATION>입니다.
- 전반기/후반기, 성수기/비수기, 환절기
  - <전반기: DT\_DURATION> 근로장려금을 신청 못 했으면 <후반기: DT\_DURATION>에 신청하면 됩니다.
  - 제주도는 <성수기: DT\_DURATION>에 가면 어딜 가도 사람이 많고 비싸니까, <비수기: DT\_DURATION>에 가는 걸 추천해요.
  - <환절기: DT\_DURATION> 건강 꿀조합템을 소개합니다.
- 하루, 이틀, 주말, 일주일, ~주, ~개월, ~한 해, ~년
  - <하루: DT\_DURATION>에 빠지는 머리카락이 너무 많다면 탈모를 의심해 봐야 합니다.
  - <이번 주말: DT\_DURATION>은 기필코 일찍 일어나서 시간을 길게 보내겠다고 다짐해 놓고, 눈 떠보니 11시였다.
  - 이날 방송에서는 '최애' 멤버들이 <일주일: DT\_DURATION>이라는 짧은 연습 기간에도 팀워크를 맞춰가며 노력하는 모습이 안방극장에 훈훈한 감동을 전했다.
  - 단양 도담삼봉은 <한 해: DT\_DURATION> 466만 명이 찾은 충청권 최고의 인기 관광지이다.
  - 오늘은 너무나 유명한 우리 전홍철 선생님을 믿고 <3년: DT\_DURATION>을 투자하신 선배의 후기를 들어 봅시다. (3년의 기간 동안을 투자한 것이므로 DT\_YEAR가 아니라 DT\_DURATION으로 태깅한다.)
- ~시즌
  - 2018 <포스트시즌: DT\_DURATION>은 두산과 SK의 맞대결에서 SK의 우승으로 끝이 났다.
  - 올해, 가장 뜨거운 매물로 평가받는 푸홀스와 필더를 비롯한 25명의 거물급 선수들이 대거 포함된 역대급 <스토브리그: DT\_DURATION>가 온다.
- ~째
  - 여행 <3일 째: DT\_DURATION> 비가 내렸다. (3일 동안 비가 내렸다는 의미이기 때문에 QT\_ORDER가 아니라 DT\_DURATION으로 태깅한다.)

- 동생이 가출한 지 <일주일째: DT\_DURATION>다.

#### 8.1.4. 주의 사항

- (D+n) 'D+6'과 같은 구성은 DT\_DURATION이 아니라 QT\_ORDER로 태깅한다.
  - 초등학교 1학년 D+6, 오늘은 남산 나들이를 했습니다. → 초등학교 1학년 <D+6: QT\_ORDER>, 오늘은 남산 나들이를 했습니다.(초등학생이 된 지 여섯 번째 날에 남산 나들이를 했다고 해석되기 때문.)
- (하루) '하루'는 '한 낮과 한 밤이 지나는 동안'을 의미하는 맥락이라면 DT\_DURATION으로 태깅한다. ('막연히 지칭하는 어떤 날'을 의미하는 맥락이라면 DT\_DAY로 태깅한다.)
  - 시골 할머니 댁에서 하루를 보냈다. → 시골 할머니 댁에서 <하루: DT\_DURATION>를 보냈다.
  - 하루는 집에 와 보니 엄마가 없었다. → <하루: DT\_DAY>는 집에 와 보니 엄마가 없었다.
- (평일, 주말) '평일'은 우리말샘 정의에 따르면 '토요일, 일요일, 공휴일이 아닌 보통 날이며 '주말'은 '한 주일의 끝 무렵, 주로 토요일부터 일요일까지'를 이른다. 이에 따라 '평일'은 DT\_DAY로 태깅하지만, '주말'은 DT\_DAY가 아니라 DT\_DURATION으로 태깅한다.
- (~주, ~개월) '~주'는 맥락에 따라 DT\_WEEK나 DT\_DURATION으로 태깅하고, '~개월'은 DT\_MONTH가 아니라 DT\_DURATION으로 태깅한다.
  - 특히 바지를 입었을 때 엉덩이가 터지기 직전이었는데, 3주가 지난 지금은 툭 떨어지는 핏으로 바뀌었습니다. → 특히 바지를 입었을 때 엉덩이가 터지기 직전이었는데, <3주: DT\_DURATION>가 지난 지금은 툭 떨어지는 핏으로 바뀌었습니다.
  - 국토교통부 공식 SNS에 "엄마아빠 신혼 때는 6개월에 한 번씩은 이사 다녔다"는 내용의 만화가 논란이 되고 있다. → 국토교통부 공식 SNS에 "엄마아빠 신혼 때는 <6개월: DT\_DURATION>에 한 번씩은 이사 다녔다"는 내용의 만화가 논란이 되고 있다.
- (겨울방학) 무순시 신화소학교는 지난 겨울방학간, 학생수공작품 전람을 진행했다. → 무순시 신화소학교는 <지난 겨울방학간: DT\_DURATION> 학생수공작품 전람을 진행했다. → '겨울방학간'도 태깅한다.
- (1학기) 'N학기'가 기간을 의미하는 것이라면 DATE, 학기를 세는 단위라면 QT로 태깅한다. 전자의 경우에는 DT\_DURATION으로 태깅한다.
  - 1학기가 끝나가는 지금, 학년별로 배운 내용과 학습 점검하는 팁을 알려드릴게요. → <1학기: DT\_DURATION>가 끝나가는 지금, 학년별로 배운 내용과 학습 점검하는 팁을 알려드릴게요.
- (1박 2일) '1박 2일'의 경우 이틀이라는 구체적인 기간을 의미하므로 DT\_DURATION으로 태깅한다.
  - 여름에 휴가를 가고 싶은 이들에게 1박 2일로 가볍게 다녀오기 좋은 속초 여행지를 소개하고자 한다. → 여름에 휴가는 가고 싶은 이들에게 <1박 2일: DT\_DURATION>로 가



법게 다녀오기 좋은 속초 여행지를 소개하고자 한다.

- (연말) '연말'은 '정초'의 반의어이지만, 우리말샘 정의(한해의 마지막 무렵)에 따라 DT\_DURATION으로 태깅한다. '연말'과 유사한 표현('연시' 등) 역시 DT\_DURATION으로 태깅한다. '연말'을 태깅 대상으로 삼은 것과 같은 맥락으로, '월초, 월말, 연초'도 태깅 대상으로 삼는다.
  - 연말이면 설레고 희망적인 낙천적인 분위기가 떠오른다. → <연말: DT\_DURATION>이면 설레고 희망적인 낙천적인 분위기가 떠오른다.
  - 벌써 월말, 그런데 월초면 없다. → 벌써 <월말: DT\_DURATION>, 그런데 월초<DT: DURATION>면 없다.
  - 마침 연초에 레프레시하기 좋은 겨울철 여행지를 추천해 드릴까 해요. → 마침 <연초: DT\_DURATION>에 레프레시하기 좋은 겨울철 여행지를 추천해 드릴까 해요.
- (연휴) '연휴'의 경우, DT\_DAY에서 '공휴일'을 태깅 대상으로 삼은 것과 같이, '연휴' 또한 태깅 대상으로 삼는다. 구체적인 날짜와 함께 결합했다면 DT\_DURATION으로, 그렇지 않으면 DT\_OTHERS로 태깅한다.
  - 오는 5월 연휴에는 대만으로 여행을 가 보려고 해요. → <오는 5월 연휴: DT\_DURATION>에는 대만으로 여행을 가 보려고 해요.
  - 요즘 일이 너무 많아서 연휴도 강제 반납하고 집에서 일만 했지 뭐야. → 요즘 일이 너무 많아서 <연휴: DT\_OTHERS>도 강제 반납하고 집에서 일만 했지 뭐야.

## 8.2. DT\_DAY

### 8.2.1. 정의 및 설명

- 특정 날짜 및 절기, 기념일(법정 기념일과 같이 특정 날짜를 알 수 있는 기념일)
- 평일, 공휴일
- 어제, 오늘, 내일, 모레 등
- 특정 날을 의미하는 단어

### 8.2.2. 분석 대상

- N일
- 요일, 평일, 공휴일
- 절기
  - 입춘, 곡우, 하지 등
- 기념일 및 명절
  - 설날, 근로자의 날, 어버이날, 빼빼로데이, 크리스마스, 수능날, 정초 등
- 어제, 오늘, 당일, 내일, 모레 등

- 놀토, 불토 등과 같이 특정 날을 의미하는 단어

### 8.2.3. 예문

- N일
  - 홍성군의회는 오는 <21일: DT\_DAY>에 제271회 후반기 첫 임시회를 개최한다.
- 요일, 평일, 공휴일
  - 일주일 중 <금요일: DT\_DAY>은 가장 마음이 편안한 날이다.
  - 대구시는 중·고등학교 방학 시행에 따라 오는 8월 3일부터 21일까지 <평일: DT\_DAY> 시내버스 52개 노선을 74대 감회 운영한다.
  - 버스가 유난히 안 온다고 했더니, 생각해 보니까 오늘이 <공휴일: DT\_DAY>이었다.
- 절기
  - <하지: DT\_DAY>는 24절기의 하나로, 북반구에서 일년 중 낮의 길이가 가장 긴 절기이다.
- 기념일 및 명절
  - 이제 며칠 후면 우리나라 최대 명절 중에 하나인 <설날: DT\_DAY>이 다가옵니다.
- 어제, 오늘, 당일, 내일, 모레 등
  - 국내 코로나19 확진자가 <어제: DT\_DAY> 하루 30명 증가했습니다.
- 놀토, 불토 등과 같이 DAY를 의미하는 단어
  - 무더운 여름 끝자락, 집에서도 <불금: DT\_DAY> 분위기를 제대로 낼 수 있는 레드불 컵 테일 레시피 5가지를 소개합니다.

### 8.2.4. 주의 사항

- (태깅 단위) DAY와 함께 구체성을 더하는 표현이 있을 경우 함께 태깅한다.
  - 지난 4일 프랑스와 중국 간의 정상회담이 열렸습니다. → <지난 4일: DT\_DAY> 프랑스와 중국 간의 정상회담이 열렸습니다.
  - 여야가 합의한 제21대 국회 개원식 개최는 16일쯤이 될 것으로 보인다. → 여야가 합의한 제21대 국회 개원식 개최는 <16일쯤: DT\_DAY>이 될 것으로 보인다.
- (보통명사) '오늘, 내일, 모레, 첫날, 마지막날, 내일, 모레, 금일, 당일'과 같은 특정 날짜를 드러내지 않는 보통명사도 태깅한다.
  - 오늘 오후면 서울 G20 정상회의는 마무리됩니다. → <오늘: DT\_DAY> 오후면 서울 G20 정상회의는 마무리됩니다.
- '매일'은 태깅하지 않는다.
- ('N주+요일' 구성) '첫째 주 화요일'과 같은 구성이 있으면 각각 따로 태깅한다.
  - 셋째 주 수요일에 만날까요? → <셋째 주: DT\_WEEK> <수요일: DT\_DAY>에 만날까요?

### 8.3. DT\_WEEK

#### 8.4.1. 정의 및 설명

- 주(週)

#### 8.4.2. 분석 대상

- 지난주, 이번 주, 다음 주 등
- 첫 주, 둘째 주 등

#### 8.4.3. 예문

- 지난주, 이번 주, 다음 주
  - <지난주: DT\_WEEK> 암 진단을 받고 식이요법을 하고 있는데 기름은 어떻게 사용해야 하나요?
- 첫 주, 둘째 주 등
  - 2일 방송된 SBS 음악 프로그램 '인기가요'에서는 8월 <첫째 주: DT\_WEEK> 1위 후보와 더불어 이날 무대에 오르는 아티스트가 소개됐다.

#### 8.4.4. 주의 사항

- (매주) '매주'는 '매일, 매달, 매월, 매해, 매년'을 태깅하지 않듯이 구체적인 표현이 뒤따른다고 하더라도 태깅하지 않는다.
- (~주) '~주'는 맥락에 따라, DURATION 또는 WEEK로 태깅되므로 주의가 필요하다.
  - BOS 비상, 헤이워드 '발목 부상'으로 '4주' 결장 예상. → BOS 비상, 헤이워드 '발목 부상'으로 '<4주: DT\_DURATION>' 결장 예상.
  - 시험 직전인 넷째 주에는 각 과목에 대한 복습과 취약점 보완을 중점으로 공부할 예정이다. → 시험 직전 <넷째 주: DT\_WEEK>에는 각 과목에 대한 복습과 취약점 보완을 중점으로 공부할 예정이다.
- (일주일) '일주일'은 월요일부터 일요일까지의 기간을 나타내는 것이므로 DURATION으로 태깅한다.
  - 일주일엔 못 되더라도 2일은 머무를 각오들을 한다고 했다. → <일주일: DT\_DURATION>은 못 되더라도 2일은 머무를 각오들을 한다고 했다.

### 8.4. DT\_MONTH

#### 8.4.1. 정의 및 설명

- 월(月)

#### 8.4.2. 분석 대상

- N월
- 이달, 지난달, 다음 달 등

#### 8.4.3. 예문

- N월
  - 제가 <9월: DT\_MONTH>에 발표했던 내용이 다시 언급되기도 했는데요.
- 이달, 지난달, 다음 달 등
  - 얼마 전, 2020년 상반기 <이달: DT\_MONTH>의 우수게임 모집공고 소식을 전해 드렸는데요.
  - 현장에서 들려오는 소식으로는 <지난달: DT\_MONTH> 경주에서 열렸던

#### 8.4.4. 주의 사항

- (N월) MONTH 앞에 '지난, 오는, 내달'과 같은 표현이 있으면 함께 태깅한다.
  - 지난 8월에 내년 파리회의 사무국 설치문제를 의제화하겠다고 프랑스가 발표했다. → <지난 8월: DT\_MONTH>에 내년 파리회의 사무국 설치문제를 의제화하겠다고 프랑스가 발표했다.
  - 영국이 내달 7월에는 사회적 거리 두기의 기준을 완화하기로 하였다. → 영국이 <내달 7월: DT\_MONTH>에는 사회적 거리 두기의 기준을 완화하기로 하였다.
- (~개월) '~개월'은 DT\_MONTH가 아니라 DT\_DURATION으로 태깅한다.
  - 그랜저는 7개월 연속 국내 최다 판매 모델로 기록됐고, 아반떼는 역대 최고 판매 기록 (2016년 6월 1만2364대) 갱신에 한층 가까워진 모양새다. → 그랜저는 <7개월: DT\_DURATION> 연속 국내 최다 판매 모델로 기록됐고, 아반떼는 역대 최고 판매 기록 (2016년 6월 1만2364대) 갱신에 한층 가까워진 모양새다.
- (초初, 말末) 의존 명사 '초, 말'이 뒤따르는 경우에도 함께 태깅한다. 중순(中旬)도 마찬가지로 함께 태깅한다.
  - 7월 말에 사무실을 옮길 예정이다. → <7월 말: DT\_MONTH>에 사무실을 옮길 예정이다.
  - 독일 정부는 5월 중순에 분데스리가 무관중 리그를 재개하기로 하였다. → 독일 정부는 <5월 중순: DT\_MONTH>에 분데스리가 무관중 리그를 재개하기로 하였다.
- (접미사) 접미사가 결합하여 의미를 구체화시키는 경우라면 접미사까지 통합하여 태깅한다.
  - 춘천시는 시민 1인당 10만 원 가량의 재난지원금을 9월쯤 지급할 것이라 발표했다. → 춘천시는 시민 1인당 10만 원 가량의 재난지원금을 <9월쯤: DT\_MONTH> 지급할 것이라 발표했다.
- '매월, 매달'은 구체적인 표현이 뒤따른다고 하더라도 태깅하지 않는다.

## 8.5. DT\_YEAR

### 8.5.1. 정의 및 설명

- 해(年)

### 8.5.2. 분석 대상

- N년(대, 경 등)
- 작년, 올해, 내년, 새해 등
- ~년생
- 육십갑자+년
  - 을미년, 병자년 등
- 십이간지+해
  - 쥐띠 해 등

### 8.5.3. 예문

- N년(대, 경 등)
  - 사피옴스에 따르면 보리 화폐는 <BC 3000년: DT\_YEAR> 수메르에서 글쓰기가 등장한 것과
  - 예를 들어서 관광산업을 보면, 환율이 제일 약했고 말하자면 외국 사람들이 제일 유리했던 것은 <2008년: DT\_YEAR>에 달러 환율이 1410원까지 올랐었어요.
  - 두 영화 가운데 '나쁜 녀석들: 포에버'는 <1990년대: DT\_YEAR>의 월 스미스를 떠올리게 만든다.
  - 이건 <2002년경: DT\_YEAR> 양수초등학교 두물머리 학교문집에 실었던 글이다.
- 작년, 올해, 내년, 새해 등
  - <올해: DT\_YEAR> 매출은 3~4월 카드랑 현금영수증 더한 금액으로 하고, <작년: DT\_YEAR> 매출을 포스기에서 총액 사진 찍어서 제출해도 되나요?
  - 여러분의 <새해: DT\_YEAR> 계획은 제대로 지켜지고 있나요?
- ~년생
- 특히 주목해야 할 선수로는 <90년생: DT\_YEAR>들의 구심점 역할을 하고 있는 21기 정하늘(종합랭킹 3위) 선수다.
- 육십갑자+년
  - 나는 지금부터 84년 전, <병자년: DT\_YEAR> 쥐띠 해에 일본 교토에서 태어났다.
- 십이간지+해
  - 나는 지금부터 84년 전, 병자년 <쥐띠 해: DT\_YEAR>에 일본 교토에서 태어났다.

### 8.5.4. 주의 사항

- (연호) '영락'과 같이 연호와 연도가 함께 있다면 태깅하지만, 연호만 있는 경우는 태깅하지 않는다.
  - 연호란 군주시대(왕)에 왕이 자신이 통치하기 시작한 해를 말하는데, 영락 1년이면 → 연호란 군주시대(왕)에 왕이 자신이 통치하기 시작한 해를 말하는데, <영락 1년: DT\_YEAR>이면
- (초初, 말末) '초, 말'과 같은 의존 명사가 뒤따르는 경우, 함께 태깅한다.
  - 박미경이라는 가수의 '민들레홀씨 되어'는 90년대 초를 대표하는 유행곡입니다. → 박미경이라는 가수의 '민들레홀씨 되어'는 <90년대 초: DT\_YEAR>를 대표하는 유행곡입니다.
- (접미사) 접미사가 결합하여 의미를 구체화시키는 경우라면 접미사까지 통합하여 태깅한다.
  - 의정부시의 지식산업센터는 2022년쯤 완공될 예정이다. → 의정부시의 지식산업센터는 <2022년쯤: DT\_YEAR> 완공될 예정이다.
- '작년, 올해, 내년, 첫해, 마지막 해' 등도 태깅한다.
- '매일, 매월, 매달'과 마찬가지로 '매년, 매해'도 태깅하지 않는다.

## 8.6. DT\_SEASON

### 8.6.1. 정의 및 설명

- 계절

### 8.6.2. 분석 대상

- 봄(철), 여름(철), 가을(철), 겨울(철)
- 춘계, 하계, 추계, 동계

### 8.6.3. 예문

- 봄(철), 여름(철), 가을(철), 겨울(철)
  - <여름: DT\_SEASON>에는 뭐니뭐니해도 평상에 앉아서 먹는 수박이 최고다.
  - <가을철: DT\_SEASON> 쌀쌀한 날씨에는 가벼운 외투를 두세 개 입는 것이 좋습니다.
- 춘계, 하계, 추계, 동계
  - 제천시는 8월 3일부터 250명의 대학생들이 참여하는 <하계: DT\_SEASON> 대학생 시간선택제 일자리사업을 실시한다고 밝혔다.

### 8.6.4. 주의 사항

- (태깅 단위) SEASON에 구체성을 더하는 표현은 함께 태깅한다.
  - 올 봄 전 세계를 덮친 중국발 코로나 바이러스 → <올 봄: DT\_SEASON> 전 세계를 덮친 중국발 코로나 바이러스

## 8.7. DT\_GEOAGE

### 8.7.1. 정의 및 설명

- 지질시대

### 8.7.2. 분석 대상

- 지질시대
  - 캄브리아기, 중생대, 원생대, 올리고세, 에오세, 신생대, 선사시대, 석탄기, 빙기, 고생대
- 구석기시대, 신석기시대, 청동기시대, 철기시대 등

### 8.7.3. 예문

- 지질시대
  - 지구의 역사에서 <캄브리아기: DT\_GEOAGE>는 무척 중요한데, 지질시대를 크게 두 부분으로 구분할 때 <캄브리아기: DT\_GEOAGE>를 기준으로 나누기 때문이다.
  - 화폐석은 <올리고세: DT\_GEOAGE>에 완전히 절멸하였고, 속씨식물이 발달하였다.
- 구석기시대, 신석기시대, 청동기시대, 철기시대
  - 한반도에 여러 국가들이 생겨나고 철로 만든 도구를 사용하게 된 시기를 <철기시대: DT\_GEOAGE>라고 한다.

### 8.7.4. 주의 사항

- (시대의 period) 시대의 기(period)가 함께 쓰인 경우, 통합하여 하나로 태깅한다.
- 고생대 오르도비스기는 고생대의 두 번째 지질시대이다. → <고생대 오르도비스기: DT\_GEOAGE>는 고생대의 두 번째 지질시대이다.
- 중생대 초기의 건조한 내륙에서 등장한 공룡은 가장 성공한 파충류로, 이후 2억년 가까이 육상을 지배하게 되었다. → <중생대 초기: DT\_GEOAGE>의 건조한 내륙에서 등장한 공룡은 가장 성공한 파충류로, 이후 2억년 가까이 육상을 지배하게 되었다.

## 8.8. DT\_DYNASTY

### 8.8.1. 정의 및 설명

- 왕조 및 그 시대

### 8.8.2. 분석 대상

- 왕조
  - 정조 시대, 가리에누스 황제 시대 등
- DT\_GEOAGE에 속하지 않는, 역사적 시대 명칭
  - 고대, 중세
  - 고구려시대, 고려시대, 조선시대, 청대, 당대 등

### 8.8.3. 예문

- 왕조
  - 정약용은 <정조 시대: DT\_DYNASTY> 청년기에 벼슬에 나가 정조의 깊은 총애를 받아 임금의 측근에 있었으나 보수 집권 세력인 노론 벽파의 공격에 시달리며 순탄치 못한 관료기를 보내게 된다.
  - 또한 <가리에누스 황제 시대: DT\_DYNASTY>에 원로원을 군무에서 축출하는 법을 통과 시키면서 군무와 정무 모두에 균형이 잡힌 인재를 배출할 수단마저 끊어졌다.
- GEOAGE에 속하지 않는 역사적 시대 명칭
  - <중세 시대: DT\_DYNASTY>에는 시학 및 음악이론에 관한 저술이 많았지만 시각예술에 대해서는 일반적인 논술이 전혀 없었다.
  - <고려시대: DT\_DYNASTY>에 들어와서는 불교가 더욱 번창하게 되고 사찰의 건립이 크게 늘어남에 따라, 다양한 종류의 불교 관계 금속공예품이 활발하게 제작되었다.

### 8.8.4. 주의 사항

- (DT\_DURATION과의 경계) '구한말, 일제강점기, 조선 시기' 등은 DT\_DURATION으로 볼 수도 있지만, 역사의 특정 시대를 나타내므로 DT\_DYNASTY로 태깅한다.
  - 덕수궁은 구한말 어지러웠던 시기에 조선의 멸망을 지켜본 궁궐이다. → 덕수궁은 <구한말: DT\_DYNASTY> 어지러웠던 시기에 조선의 멸망을 지켜본 궁궐이다.
- (~때) 왕의 이름 뒤에 '~때'가 붙으면 DT\_OTHERS가 아니라 DT\_DYNASTY로 태깅한다.
  - 조선 전기의 문물제도는 성종 때 거의 완성되었다고 할 수 있다. → <조선 전기: DT\_DYNASTY>의 문물제도는 <성종 때: DT\_DYNASTY> 거의 완성되었다고 할 수 있다.
- (고대, 중세) '현대'는 '고대'와 '중세'와 달리 시기를 확정할 수 없으므로 태깅하지 않는다.

## 8.9. DT\_OTHERS

### 8.9.1. 정의 및 설명

- DATE의 어떤 분류에도 들어가지 않는 기타 날짜
- 불명확한 기간, 주기를 나타내는(DT\_DURATION에 속하지 않는) 날짜



### 8.9.2. 분석 대상

- '년+월+일'이 함께 쓰이는 경우 (셋 중 둘만 함께 쓰이는 경우 및 영문도 포함)
- ~ 전, ~후, ~뒤, ~중, ~내 (구체적인 날짜와 쓰여도 '100년 전, 이번 주 중' 등과 같이 특정 시점을 알 수 없기 때문)
- ~에서 ~사이
- ~부터, ~까지 ( '~부터'나 '~까지'가 단독으로 사용되는 경우 구체적인 날짜와 결합했다고 하더라도, '부터, 까지'가 나오면 그 시점을 특정할 수 없기 때문)
- ~세기
- ~때 ('제 2차 세계대전 때, 임진왜란 때'와 같이 특정 사건과 '때'가 함께 쓰이는 경우)

### 8.9.3. 예문

- '년+월+일'
  - <2007년 10월 25일: DT\_OTHERS> OECD에 들어갔다.
  - <작년 10월 6일: DT\_OTHERS> 중요한 시험이 있었다.
  - <20년 6월 말: DT\_OTHERS> 기준 서해선 복선전철 건설사업 노반공사 공정을 공유드리겠습니다.
  - <18/Dec/2019: DT\_OTHERS> 아침 운동으로 요즘은 트레드밀이 근력운동에 좀 밀린 것 같아 뛰어 봤습니다.
- ~전, ~후, ~뒤, ~중, ~내
  - <100년 전: DT\_OTHERS> 일어난 살인사건은 누구에 의해, 무슨 이유에서 발생했으며, 수사는 어떻게 진행되었을까?
  - 대학 면접에서 <20년 후: DT\_OTHERS> 모습을 상상하기라는 질문이 있으면 어떻게 대답해야 할까요?
  - 코로나발 실업자가 122만 8,000명으로 <1999년 이후: DT\_OTHERS> 21년 만에 최대치를 기록했다.
  - 설레는 마음으로 준비하다 보니 <3일 뒤: DT\_OTHERS>면 출국날이에요.
  - 테슬라 모델 S가 새로운 모습으로 <이번 주 중: DT\_OTHERS> 공개된다!
  - 윌리엄은 룸메이트에게 <일주일 내: DT\_OTHERS>에 자신을 죽여달라고 부탁했다.
- ~에서 ~사이
  - 1981년부터 매년 <1월에서 2월 사이: DT\_OTHERS> 휴양 도시 다보스에서 열렸다.
- ~부터, ~까지
  - 우리 아이는 양쪽 무릎이 아프다고 해서 <그저께부터: DT\_OTHERS> 병원에 입원해 있어요.
  - <오는 31일부터: DT\_OTHERS> 기말고사가 시작된다.
  - 정부는 <2012년 전까지: DT\_OTHERS> 건물과 차량에서 배출하는 탄산가스의 양을 1999년에 비해 12.5% 줄일 것을 약속했다.

- ~세기
  - 콜럼버스는 <15세기: DT\_OTHERS>의 세계 지도를 보고 항해를 결심하게 되었다고 한다.
- ~때
  - 당시 국군은 자체 전투복이 없어 <제2차 세계대전 때: DT\_OTHERS> 미군이 입었던 군복 등을 입고 전쟁을 치렀다.

#### 8.9.4. 주의 사항

- ('연/월/일+요일'의 구성) '20/11/15'와 같은 구성은 통합하여 DT\_OTHERS로 태깅한다. 이때 '요일'이 추가된 경우, 요일은 따로 DT\_DAY로 태깅한다.
  - 행사 날짜는 2020.1.13(수)입니다. → 행사 날짜는 <2020.1.13: DT\_OTHERS>(<수: DT\_DAY>)입니다.
- (지난~) '지난'은 단독 태깅하지 않으며, '지난+DT 범주에 속하는 명사 또는 명사구'일 경우에만 태깅한다. '지난+EV'나 '지난+OG'와 헷갈리지 않도록 한다.
  - 지난 1980년 5월의 움직임 덕분에 대한민국은 민주주의 국가로 재탄생하게 된 것이다. → <지난 1980년 5월:DT\_OTHERS>의 움직임 덕분에 대한민국은 민주주의 국가로 재탄생하게 된 것이다.
  - 손흥민은 지난 아시안게임에서 금메달을 땀다. → 손흥민은 지난 <아시안게임: EV\_SPORTS>에서 금메달을 땀다.
  - 손흥민은 지난 2014년 10월 아시안게임에서 금메달을 땀다. → 손흥민은 <지난 2014년 10월: DT\_OTHERS> <아시안게임: EV\_SPORTS>에서 금메달을 땀다.
- (DT+EV+~때) '작년 아시안게임 때', '지난 2002년 한일 월드컵 때'와 같이 'DT+EV+~때'의 구성이면 통합하지 않고 각각 따로 태깅한다. 이때 'EV+~때'는 DT로 태깅한다. 즉, DT+DT가 된다.
  - 지난 2002년 한일 월드컵 때 연장전에서 안정환 선수의 골든볼은 기적처럼 느껴졌었지. → <지난 2002년: DT\_YEAR> <한일 월드컵 때: DT\_OTHERS> 연장전에서 안정환 선수의 골든볼은 기적처럼 느껴졌었지.

## 9. TIME

TIME은 시계상으로 나타나는 시/시각이나 시간 범위를 나타내는 시간 표현이다. 세부 범주로는 기간 TI\_DURATION, 시/시각 TI\_HOUR, 분 TI\_MINUTE, 초 TI\_SECOND로 세분되어 있다. 이외의 시간 표현은 TI\_OTHERS로 태깅한다.

- 날짜/시간/수량과 관련하여 의미를 구체화시키는 '간, -경, 만, -씩, 안, 약, -여, 전, 후, 중, -쯤, 부터, 까지, -께, 총, 한, 가량, 안팎, 이내, 이상, 이하, 정도' 등의 경우 태깅 범위에 포함한다.

## 9.1. TI\_DURATION

### 9.1.1. 정의 및 설명

- 기간
  - 지속되는 기간
  - 어떤 시각에서 어떤 시각까지의 사이

### 9.1.2. 분석 대상

- 지속되는 기간
  - ~ 시간, ~ 시간 동안, 분, 초
  - 오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤
- 어떤 시각에서 어떤 시각까지의 사이
  - ~부터 ~까지

### 9.1.3. 예문

- ~ 시간, ~ 시간 동안, 분, 초
  - 이들은 피해자를 <다섯 시간:TI\_DURATION> 넘게 끌고 다녔다.
  - 하루 <24시간: TI\_DURATION>
  - 갓 낳은 새끼는 연한 노란색으로 생후 <이십 분: TI\_DURATION>이면 달릴 수 있다.
  - 김해국제공항까지의 거리가 <약 20분: TI\_DURATION> 단축되었다.
  - 물을 넣고 <5분: TI\_DURATION>
  - 300m를 주파하는데 걸린 시간은 단 <15초: TI\_DURATION>입니다.
  - 평균 액세스 타임도 <0.5초 정도: TI\_DURATION> 빠르다.
- 오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤
  - 원장님은 <오전: TI\_DURATION>에만 환자를 본다.
  - 나쁜 봄날 <오후: TI\_DURATION>
  - 어제부터 비가 오더니 <점심: TI\_DURATION>에 그쳤다.
  - 평화로운 <저녁: TI\_DURATION> 시간
  - 이른 <새벽: TI\_DURATION>이라서 오가는 사람이 별로 없다.
  - <낮: TI\_DURATION>에 낮잠을 잤더니 <밤: TI\_DURATION>에 잠이 오지 않는다.
  - <밤: TI\_DURATION>에 잠이 안 올 때 따뜻한 우유나 차를 마시고 자면 꼭 잘 수 있다.
- ~부터 ~까지
  - <3시부터 5시까지: TI\_DURATION> 브레이크 타임입니다.
  - 영업 시간: 매일 <10:30 ~ 21:00: TI\_DURATION>
  - 금요일은 <오후 4시부터 밤 8시까지: TI\_DURATION>입니다.

#### 9.1.4. 주의 사항

- 기간이나 지속 시간 등은 모두 TI\_DURATION에 해당한다.
- '3시간 연속'에서 '연속'은 태깅하지 않는다.
- '전, 후, 이전, 이후, 직전, 직후' 등은 DT, TI, EV에 한정해서 태깅한다.
  - 강의 시간 <10분 전: TI\_OHTERS>에 도착은 예의라고 꼭꼭 적어 두었다.
- '오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤'이 시간과 함께 쓰여 시간을 나타낼 때는 묶어서 태깅한다.
  - 금요일은 오후 4시부터 밤 8시까지입니다. → 금요일은 <오후 4시부터 밤 8시까지: TI\_DURATION>입니다.
- 부사 '밤낮없이'는 태깅 대상이 아니다.
  - 이틀동안 밤낮없이 비가 내렸다.
- 시간을 나타내는 개체명 내부의 기호(~, +, -, :)는 포함하여 태깅한다.
  - 2+2 총 4시간 → <2+2: TI\_DURATION> <총 4시간: TI\_DURATION>
- '시간, 분, 초'가 문맥상 걸린 시간을 나타낼 때는 TI\_DURATION으로 태깅한다.
  - 학생들이 인터뷰만 3시간째다. → 학생들이 인터뷰만 <3시간째: TI\_DURATION>다.
  - 사진을 고르는 데 15분 → <15분: TI\_DURATION>
  - 진동이 3초 정도 계속되었다. → 진동이 <3초 정도: TI\_DURATION> 계속되었다.
- '낮잠', '어젯밤', '여름밤'과 같은 복합어는 내부 태깅하지 않는다.
- 끼니로 먹는 음식을 나타내는 '아침, 점심, 저녁'은 태깅 대상이 아니다.

## 9.2. TI\_HOUR

### 9.2.1. 정의 및 설명

- 시간의 어느 한 시점, 하루의 24분의 1이 되는 동안을 세는 단위
  - 시간 및 시각

### 9.2.2. 분석 대상

- 시간 및 시각
  - 시

### 9.2.3. 예문

- 시
  - 수험생들은 오늘 오전 재학 중인 학교 등에서 시험장이 기재된 수험표를 배부받은 뒤 <오후 3시: TI\_HOUR> 시험장으로 찾아가서 예비 소집에 참가해야 합니다.
  - 마르코스 씨는 <여섯 시: TI\_HOUR>에 집에 도착했습니다.

- 지금 시각은 <오후 2시: TI\_HOUR>입니다.

#### 9.2.4. 주의 사항

- 특정 시각만 TI\_HOUR에 해당한다. 지속되는 시간은 TI\_DURATION으로 태깅하고 '특정 시각+전, 후, 부터, 까지'를 나타낼 때는 TI\_OTHERS로 태깅하여 특정 시각만 해당하는 TI\_HOUR의 태깅과 구별하여야 한다.
- 오전, 오후는 시간과 함께 태깅한다.
  - 그 백화점의 개점 시간은 오전 10시이다. → 그 백화점의 개점 시간은 <오전 10시: TI\_HOUR>이다.

### 9.3. TI\_MINUTE

#### 9.3.1. 정의 및 설명

- 한 시간의 60분의 1이 되는 동안을 세는 단위
  - 분, 반(30분)

#### 9.3.2. 분석 대상

- 분, 반(30분)

#### 9.3.3. 예문

- 분, 반(30분)
  - 지금 시간이 2010년 마지막 11시 56분, <57분: TI\_MINUTE>을 지나가고 있습니다.
  - A: 지금 몇 분이야? B: <30분: TI\_MINUTE>!

#### 9.3.4. 주의 사항

- 특정 시각의 분 단위를 말할 때만 TI\_MINUTE에 해당한다. 지속되는 시간은 TI\_DURATION으로 태깅하고 '특정 시각 분+전, 후(예: 영화 시작 30분 전에 도착했어요.), 부터, 까지'를 나타낼 때는 TI\_OTHERS로 태깅한다. 특정 시각을 나타내는 시간 단위인 분만 TI\_MINUTE으로 태깅해야 한다.

### 9.4. TI\_SECOND

#### 9.4.1. 정의 및 설명

- 한 시간의 3,600분의 1이 되는 동안을 세는 단위

- 초

#### 9.4.2. 분석 대상

- 초

#### 9.4.3. 예문

- 초
  - <1초: TI\_SECOND>, <15초: TI\_SECOND>, <30초: TI\_SECOND>
  - 현재 시각은 11시 59분 55초, <56초: TI\_SECOND>를 막 지나고 있습니다.

#### 9.4.4. 주의 사항

- 특정 시각의 초 단위 시간을 말할 때만 TI\_SECOND에 해당한다. 지속되는 시간을 나타내는 초 단위는 TI\_DURATION으로 태깅하고 '특정 시각 초 단위+전, 후, 부터, 까지'를 나타낼 때는 TI\_OTHERS로 태깅한다.

### 9.5. TI\_OTHERS

#### 9.5.1. 정의 및 설명

- TIME의 세부 분류가 아닌 기타 시간

#### 9.5.2. 분석 대상

- TIME의 세부 분류가 아닌 기타 시간
  - 시+분+초, 시+분, 분+초
  - 00:00:00
  - 시간/특정 시각(시, 분, 초)+전, 후, 이전, 이후, 부터, 까지  
( '~부터'나 '~까지'가 단독으로 사용되는 경우, 특정 시각(시, 분, 초)과 결합했다고 하더라도 '부터, 까지'가 나오면 그 시점을 특정할 수 없기 때문)

#### 9.5.3. 예문

- 시+분+초, 시+분, 분+초
  - 지금 시간이 2010년 마지막 <11시 56분 55초: TI\_OTHERS>, 56초를 지나가고 있습니다.
  - 밍키님과 <16시 20분: TI\_OTHERS>에 만나기로 했다.
  - <12시 반: TI\_OTHERS>에 가 봄.
- 00:00:00
  - 한국 남자 마라톤 신기록은 이봉주 선수가 세운 <02:07:20: TI\_OTHERS>이다.

- <12:00: TI\_OTHERS>에 출발
- 시간/특정 시각(시, 분, 초)+전, 후, 이전, 이후, 부터, 까지
  - 러시아에서 발사되었으며 <6시간 후: TI\_OTHERS> 지상국과 교신에 성공하였다.
  - <오전 10시 이전까지: TI\_OTHERS> 주문시 당일 배송됩니다.
  - 표를 사기 위해 <새벽 5시부터: TI\_OTHERS> 줄을 서기 시작했다.
  - <아침부터: TI\_OTHERS> 비가 내린다.
  - 영화 시작 <30분 전: TI\_OTHERS>에 도착했어요.
  - 로켓 발사 <10초 전: TI\_OTHERS>

#### 9.5.4. 주의 사항

- '시+분+초, 시+분, 분+초'가 함께 쓰일 경우에는 전체를 묶어서 TI\_OTHERS로 태깅한다.
  - 현재 시각은 11시 59분 57초입니다. → 현재 시각은 <11시 59분 57초: TI\_OTHERS>입니다.
- (스포츠 기록)스포츠의 기록을 나타내기 위해 '시+분+초, 시+분, 분+초' 등이 함께 쓰인 '02:07:20, 03'15", 05:37'과 같은 시간 표현은 TI\_OTHERS로 태깅한다.
  - 한국 남자 마라톤 신기록은 이봉주 선수가 세운 <02:07:20: TI\_OTHERS>이다.
- '12:30'과 같이 시간을 나타내는 경우는 전체를 묶어서 TI\_OTHERS로 태깅한다.
  - 13:00시 정각에 모인다. → <13:00시: TI\_OTHERS> 정각에 모인다.
- '전, 후, 이전, 이후, 직전, 직후' 등은 DT, TI, EV에 한정해서 태깅한다.
- '오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤'이 시간이나 특정 시각을 나타내는 단위와 함께 쓰였을 때는 묶어서 태깅한다.
  - 오후 18:30에 저녁을 먹는다. → <오후 18:30: TI\_OTHERS>에 저녁을 먹는다.
  - 청담동 애플떡볶이를 먹기 위해 오전 9시부터 기다렸다. → 청담동 애플떡볶이를 먹기 위해 <오전 9시부터: TI\_OTHERS> 기다렸다.

## 10. QUANTITY

- 수사 또는 수관형사와 함께 쓰여 세거나 잴 수 있는 양(수량, 분량) 표현
- 수사 또는 수관형사와 함께 쓰여 순서, 순차를 나타내는 표현
- 기타 수사로 이루어진 표현
- 주의 사항
  - 수량과 함께 쓰여 의미를 구체화시키는 '사이, -경, 만, -씩, 약, -여, -쯤, 부터, 까지, -께, 총, 한, 가량, 안팎, 이내, 이상, 이하, 정도, -밖에, -뿐' 등을 포함하여 태깅한다.
  - 'n1부터 n2까지', 'n1~n2'와 같이 범위를 나타낼 경우 'n1'의 분류 기준에 따라 전체를

태깅한다.

- pp. 55~67: p.55가 QT\_ORDER이기 때문에 <pp.55~67: QT\_ORDER>와 같이 태깅한다.
- 4세부터 7세까지: '4세'가 QT\_AGE이기 때문에 <4세부터 7세까지: QT\_AGE>와 같이 태깅한다.
- SIZE/LENGTH/WEIGHT/TEMPURATURE/VOLUME/SPEED/PERCENTAGE는 실제로 사용되는 단위 표지를 기준으로 태깅한다.
  - 각 항목의 "분석대상"을 참고할 것.
  - (예외) 신발 사이즈 등과 같이 크기와 관련된 표현에 길이의 단위를 사용할 경우에는 QT\_LENGTH로, 강설량, 강우량, 강수량과 같이 길이의 단위를 써서 부피나 용량을 나타내는 경우에는 QT\_VOLUME으로 태깅한다. (각각 지침 10.3, 10.10 참조)
  - (두께)는 사물의 한 면과 마주한 다른 한 면 사이의 거리를 측정한 값이므로 거리(길이)의 단위를 써서 QT\_LENGTH로 태깅한다. (지침 10.3.4 참조)
- DT와의 구분: 이하의 기준에 따라 구분하여 태깅한다.
  - [n일] '일'을 단위로 날의 수를 세는 경우는 QT\_COUNT, '얼마 동안'의 경우는 DT\_DURATION, 달력 상의 날짜는 DT\_DAY로 태깅한다.
    - 올해 공휴일은 31일밖에 안 된다. → 올해 공휴일은 <31일밖에: QT\_COUNT> 안 된다.
    - 커플 31일 기념 프사 → 커플 <31일: DT\_DURATION> 기념 프사 ('31일'은 31일 동안 만난 것을 기념하는 의미이므로 QT\_ORDER가 아니라 DT\_DURATION으로 태깅)
    - 31일 5시 출발 → <31일: DT\_DAY> 5시 출발
  - [n일째] '동안'의 의미로 쓰였으면 DT\_DURATION, '순차'의 의미를 나타내면 QT\_ORDER로 태깅한다.
  - 3일째 되는 날 비가 내리기 시작했다. → <3일째: QT\_ORDER> 되는 날 비가 내리기 시작했다.
  - 3일째 비가 내렸다. → <3일째: DT\_DURATION> 비가 내렸다.
  - [n일차] 日次가 '날의 차례'를 의미하므로 DT가 아닌 QT로 태깅한다. (비슷한 쓰임의 '年次'에 대한 우리말샘 해설을 보면, 그 상위어가 '차례'라고 되어 있음.)
    - 교생실습 1일차 → 교생실습 <1일차: QT\_ORDER>
    - 입사 8년차 → 입사 <8년차: QT\_ORDER>

## 10.1. QT\_AGE

### 10.1.1. 정의 및 설명



- 사람이나 동·식물 등이 태어나 살아온 햇수(연령), 개월수(월령), 주일수[(생후) n주], 일수[(생후) n일]
- 하위 유형
  - 사람의 나이
  - 동물의 나이
  - 식물의 나이

### 10.1.2. 분석 대상

- 사람의 나이
  - 만 55세, 45세, 세 살 등
  - 8개월, 14개월령 등
  - 생후 10주, 생후 180일 등
  - 10대, 50~60대, 18~41세 등
- 동물의 나이
  - 한 살, 열 살, 11살, 1~2살 등
  - 10개월, 3개월령 등
  - 3주, 50일, 7일령 등
- 식물의 나이
  - 500살, 50년 등
  - 10~15살 등

### 10.1.3. 예문

- 사람의 나이
  - 어느 광산촌에서 아버지마저 잃고 가난하게 살아가던 <10세: QT\_AGE> 소녀 안말숙.
  - <세 살: QT\_AGE> 버릇, 허리 디스크의 원인이 된다.
  - 현재 <18개월: QT\_AGE> 남아인데 아이챌린지 월령프로그램을 해도 괜찮을까요?
  - 시환이 <생후 228일: QT\_AGE>
  - (생후) <24개월 11일: QT\_AGE> 된 쌍둥이들
  - <5주: QT\_AGE> 된 영아 유기 사건
  - <10대: QT\_AGE> 청소년을 연령단계로 나눠 <16세 이하: QT\_AGE>를 로틴(low teen), <17세 이상: QT\_AGE>을 하이틴(high teen)이라고 부르기도 한다.
  - <1~2살: QT\_AGE> 아기들은 입으로 탐색하는 '구강기'에 해당하며, 이 시기는 작은 물체를 반복적으로 집으며 소근육 발달이 이루어지는 시기입니다.
  - 이 전시회는 <14세 이상부터: QT\_AGE> 관람 가능합니다.
- 동물의 나이
  - 우리집 고양이 다다는 <다섯 살: QT\_AGE> , 호야는 <네 살: QT\_AGE>인데, 사람 나이

로 환산하면 각각 <37세: QT\_AGE>, <33세: QT\_AGE>이다.

- 나의 첫 고양이 미미는 <16살까지: QT\_AGE> 살았다.
- <0~6개월: QT\_AGE>은 로얄캐닌 베이비캐트를, <6개월~1살: QT\_AGE>의 경우에는 로얄캐닌 키튼을 급여한다.
- 포획 대상은 <4개월령 이상: QT\_AGE>의 고양이이며 금년에 선정되지 않은 지역은 내년도 사업 시행 시 우선 포획지로 선정할 예정이다.
- <생후 4주: QT\_AGE> 된 아기 고양이
- 종계병아리 <21일령: QT\_AGE> 모습
- 식물의 나이
  - 가장 오래된 것은 미국 캘리포니아 주 부근의 산에서 자라는 브리슬콘 소나무이구요. 나무의 나이가 <5,000살: QT\_AGE>이 넘는 것도 있대요.

#### 10.1.4. 주의 사항

- '-살', '-세' 등의 단위가 없어도 맥락 상 나이를 의미하면 QT\_AGE로 태깅한다.
  - 내 나이 24. → 내 나이 <24: QT\_AGE>.
  - 고양이 나이 20이면 우리 할머니랑 거의 동급이야. → 고양이 나이 <20: QT\_AGE>이면 우리 할머니랑 거의 동급이야.
- 일령과 관련하여, 백일(百日)이 '행사'나 '기간'의 의미를 가질 때에는 QT로 태깅하지 않는다.
  - 시환이 백일에 와 주셔서 감사합니다. → 시환이 백일에 와 주셔서 감사합니다. ('백일잔치'의 축소지칭인 '백일'은 '행사'의 의미이므로 태깅하지 않는다.)
  - 우리 아가 100일 사진 → 우리 아가 <100일: DT\_DURATION> 사진 (여기서의 '100일'은 100일이라는 기간을 기념하는 의미이므로 QT\_AGE가 아니라 DT\_DURATION으로 태깅한다.)
  - (생후) 100일 된 아기 → (생후) <100일: QT\_AGE> 된 아기 (일령, 즉 명확하게 나이를 의미할 때에만 QT\_AGE로 태깅한다.)
- 생후 n일, n주와 관련하여, 아래와 같이 문맥상 'after one's birth'를 전제하는 것으로 판단되면 QT\_AGE로 태깅한다.
  - 남자의 육아, 꼬물이 김시환 220days → 남자의 육아, 꼬물이 김시환 <220days: QT\_AGE>
  - 우리 지우 150일, 뒤집기 성공! → 우리 지우 <150일: QT\_AGE>, 뒤집기 성공!
  - 심장에 해로운 4주 애기냥 → 심장에 해로운 <4주: QT\_AGE> 애기냥
- [D+n(일)] 의 구성(QT\_ORDER)이 "생후 n일"을 의미할 때에는 QT\_AGE로 태깅한다.
  - 빈나네 마당 냥이 4남매 D+100!! 백일잔치~츄르잔치~ → 빈나네 마당 냥이 4남매 <D+100: QT\_AGE> 백일잔치~츄르잔치~ (문맥상 생후 n일째의 의미이므로 나이로 해석)

- 초등학교 1학년 D+6, 오늘은 남산 나들이를 했습니다. → 초등학교 1학년 <D+6: QT\_ORDER>, 오늘은 남산 나들이를 했습니다.(초등학생이 된 지 여섯 번째 날에 남산 나들이를 한 것이므로 순차로 해석)

## 10.2. QT\_SIZE

### 10.2.1. 정의 및 설명

- 일정한 평면에 걸쳐 있는 공간이나 범위의 크기
- 하위 유형
  - 넓이, 면적

### 10.2.2. 분석 대상

- 넓이, 면적
  - 1제곱미터(m²), 1아르(a), 1헥타르(ha), 9,606.33제곱킬로미터(km²), 10.76391제곱피트(ft²), 1.19599제곱야드(yd²), 247에이커(ac), 10.89평방자, 25평, 800단보, 20만 정보, 1m², 1a, 1000ha, 6만 8,000 km², 28km², 100yd², 247ac, 20만ac, 7천 yd², 7ft², 1경(頃)

### 10.2.3. 예문

- 넓이, 면적
  - 면적은 <33.3 km²: QT\_SIZE>, 인구는 19만 3,214명이다.
  - 보통 <1 제곱미터: QT\_SIZE>(m²)의 땅, 그러니까 이 정도 넓이의 땅에는 200마리 정도의 지렁이가 있는데요.
  - <1ha: QT\_SIZE>(<3000평: QT\_SIZE>) 기준 연간 소득은 곰취(시설재배)가 6,300만 원으로 가장 높았다.
  - <1경: QT\_SIZE>은 <3000평: QT\_SIZE>으로 <약 9,917.4m²: QT\_SIZE>에 해당한다.

### 10.2.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 1제곱미터(m²) → <1 제곱미터: QT\_SIZE>(m²)
  - 1경(頃) → <1경: QT\_SIZE>(頃)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 1ha(3,000평) → <1ha: QT\_SIZE>(<3,000평: QT\_SIZE>)
  - 1경(약 9,917.4m²) → <1경: QT\_SIZE>(<약 9,917.4m²: QT\_SIZE>)

## 10.3. QT\_LENGTH

### 10.3.1. 정의 및 설명

- 한쪽 끝에서 다른 쪽 끝까지의 거리
- 하위 유형
  - 길이, 거리, 높이

### 10.3.2. 분석 대상

- 길이, 거리, 높이
  - 10밀리미터(mm), 1센티미터(cm), 0.01미터(m), 0.00001킬로미터(km), 0.393701인치(in), 0.032808피트(ft), 0.010936야드(yd), 6.2137e-6마일(mile), 0.033자[尺(척)], 0.0055간(間), 0.000092정(町), 0.000025리(里), 5.3996e-6해리(海里), 가로 54mm, 900m, 8km, 82 피트, 7cm, 6~12cm, 3~5m, 2,200km

### 10.3.3. 예문

- 길이
  - 정아의 발 사이즈는 <240mm: QT\_LENGTH>이다.
  - 앞머리를 <2cm정도만: QT\_LENGTH> 잘라 주세요.
- 거리
  - 가항(可航) 거리는 상류에서 <700km: QT\_LENGTH>, 하류에서 <300km: QT\_LENGTH>이며, 중류는 항행이 불가능하다.
  - 르커저시(日喀則市)와의 거리는 <약 320km: QT\_LENGTH>이다.
  - 1852년에는 두 핵 사이의 거리가 <2.5106km: QT\_LENGTH>인 쌍혜성이 출현하였다.
- 높이
  - 그는 키가 <2m: QT\_LENGTH>가 넘는다.
  - J 자형의 섬으로 <높이 2,987m: QT\_LENGTH>에 달하는 산들이 솟아 있으며
  - 해발고도 <3,104m: QT\_LENGTH>의 고지에 자리 잡고 있다.

### 10.3.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 2미터(m) → <2미터: QT\_LENGTH>(m)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 1mile(1,760yd) → <1mile: QT\_LENGTH>(<1,760yd: QT\_LENGTH>)
- (두께) 어떤 한 면과 마주한 다른 면 사이의 거리를 측정한 값이므로 QT\_LENGTH로 태깅한다.
  - 이 삼겹살은 두께가 10cm는 되겠다. → 이 삼겹살은 두께가 <10cm: QT\_LENGTH>는 되겠다.

- 이 책의 세네카(책등 두께)는 105mm이다. → 이 책의 세네카(책등 두께)는 <105mm: QT\_LENGTH>이다.

## 10.4. QT\_COUNT

### 10.4.1. 정의 및 설명

- 사물의 수량이나 빈도 등을 세는 표현
  - 수량
  - 빈도
  - [D-n(일)]

### 10.4.2. 분석 대상

- 수량
  - 6개, 8백 70선, 60 개국, 1 마리, 두 마리, 약 600 종, 약 3 만 종, 약 1,000 종, 9~10 마리, 14 종, 10 권, 9 개, 7 가지, 여섯 가지, 5,000 자, 4,123 석, 390 여 칸, 18 행, 13 척, 1,600 여 개, 한두 가지, 대여섯 채, 한 학기, 하나, 둘, 세 끼, 한 잔, 석 잔(3잔), TOP 10, 3대, 5대(大) 등
- 빈도
  - 두 차례, 세 차례, 40여 차례, 97건, 77만 9472건 등
  - 이중, 4중, 한 겹, 100겹 등
- [D-n(일)]
  - D-100일, D-3

### 10.4.3. 예문

- 수량
  - 거머리들의 몸은 정확히 <34개: QT\_COUNT>의 환절로 이루어져 있는데 그 이상도 그 이하도 아니죠.
  - 어떤 거머리들의 입은 이렇게 <세 갈래: QT\_COUNT>로 갈라지는데
  - 한 쪽마다 <100개: QT\_COUNT>도 넘는 이빨이 나 있어서 마치 톱처럼 먹이의 피부를 자를 수 있죠.
  - 버섯 종류 <한두 가지 정도만: QT\_COUNT> 넣어 주셔두 괜찮아요.
  - 오늘은 너무 바빠서 <두 끼밖에: QT\_COUNT> 못 먹었어.
  - 00여대 예술계 학생의 <한 학기: QT\_COUNT> 등록금은 500만 원이 넘는다.
  - 상반기 전자제품 <베스트 10: QT\_COUNT>을 발표하겠습니다.
  - 세계 <7대: QT\_COUNT> 불가사의!!

- 빈도
  - 태풍 피해 신고 전화가 벌써 <이십 건 이상: QT\_COUNT> 걸려 왔다.
  - 사장님에게 오늘만 <두 차례: QT\_COUNT> 지적을 받았다.
  - 어둠 속으로부터 소리가 그것도 <이중: QT\_COUNT>, <삼중: QT\_COUNT>으로 겹쳐 들려오는 소리가 마치 단단한 물질처럼 우리를 뺨 둘러쌌다.
  - <세 겹: QT\_COUNT>으로 싹 꾸러미.
- [D-n(일)]
  - 강원 평창군 용평리조트와 발왕산 일대에서 개최된 2021 평창평화포럼 <D-100일: QT\_COUNT> 봄엽 평화 릴레이 걷기행사가 3일 성황리에 마무리됐다.
  - 수능 <D-3: QT\_COUNT>!!!

#### 10.4.4. 주의 사항

- 오류, 십계명과 같은 경우는 태깅하지 않는다.
- 질량수는 QT\_COUNT로 태깅한다.
  - 질량수 79 - 존재비율 50% 질량수 81 - 존재비율 50% → <질량수 79: QT\_COUNT> - 존재비율 50% <질량수 81: QT\_COUNT> - 존재비율 50%
- 'n+[사람]을 제외한 보통명사'는 모두 QT\_COUNT로 태깅한다.
  - 짜장면 한 그릇을 친구와 나눠 먹었다. → 짜장면 <한 그릇: QT\_COUNT>을 친구와 나눠 먹었다.
  - 세 과목 평균 점수가 80 이하인 경우 불합격이다. → <세 과목: QT\_COUNT> 평균 점수가 80 이하인 경우 불합격이다.
  - 사람이 많아서 두 차에 나눠 탔다. → 사람이 많아서 <두 차: QT\_COUNT>에 나눠 탔다.
- [부사] '한번'은 태깅 대상이 아니다. (띄어쓰지 않음/ "두 번", "세 번" 등으로 교체해서 어색하면 부사)
  - '언제 밥 한번 먹자.', '제가 한번 해볼게요.', '한번은 이런 일도 있었다니까요.', '그 녀석 목소리 한번 크네.' → 전부 태깅하지 않음.
- [한잔하다] "간단하게 한 차례 차나 술 따위를 마시다."라는 뜻의 동사이므로 QT로 태깅하지 않는다.
  - 다이아 정채연·기희현 술 한잔하다 나온 트로트곡 '꽃, 달, 술' → 태깅하지 않음.
  - 커피 한잔할까? → 태깅하지 않음.
  - 감식초 두 잔 할래, 까나리 한 잔 할래? → 감식초 <두 잔: QT\_COUNT> 할래, 까나리 <한 잔: QT\_COUNT> 할래?
- [n일] '일'을 단위로 날의 수를 세는 경우는 QT\_COUNT, '얼마 동안'의 경우는 DT\_DURATION, 달력 상의 날짜는 DT\_DAY로 태깅한다.
  - 올해 공휴일은 31일밖에 안 된다. → 올해 공휴일은 <31일밖에: QT\_COUNT> 안 된다.
  - 커플 31일 기념 프사 → 커플 <31일: DT\_DURATION> 기념 프사 ('31일'은 31일 동안)

만난 것을 기념하는 의미이므로 QT\_ORDER가 아니라 DT\_DURATION으로 태깅한다.)

- 31일 5시 출발 → <31일: DT\_DAY> 5시 출발

## 10.5. QT\_MAN\_COUNT

### 10.5.1. 정의 및 설명

- 사람의 수를 세는 표현
  - 인원수: n+명, 인, 분 등
  - n+보통명사(사람)
  - n+인승, 인용, 인분, 인방 등

### 10.5.2. 분석 대상

- 인원수
  - 50명, 1명, 약 600명, 12인, 2억 1,219만 5,000명, 한 명, 수천 명 등
  - 대여섯 사람, 다섯 타자, 세 학생 등
  - 4인승, 2인용, 3인분, 3인방, 오인방 등

### 10.5.3. 예문

- 인원수
  - 경찰은 주지훈씨 외에도 <10여 명: QT\_MAN\_COUNT>의 마약투약 혐의를 포착하고 혐의를 입증하는 데 주력하고 있어 연예계 마약파문은 당분간 계속 이어질 것 같습니다.
  - 많은 스태프들에게 둘러싸여 충실력을 뽐내고 있는 <여섯 명: QT\_MAN\_COUNT>의 스타들.
  - 이 쪽에 <한두 명만: QT\_MAN\_COUNT> 더 와 봐.
  - 지나가던 <세 학생: QT\_MAN\_COUNT>의 도움으로 위기를 모면할 수 있었다.
  - 다음은 덕촌리 꽃할매 <오인방: QT\_MAN\_COUNT>입니다.
  - 떡볶이 <5인분: QT\_MAN\_COUNT>을 셋이서 먹었다.
  - 현재 롤스로이스가 판매하고 있는 컨버터블 '단'은 <4인승: QT\_MAN\_COUNT> 모델이다.
  - 디자인은 물론 실용성까지 갖춘 <3인용: QT\_MAN\_COUNT> 쇼파 '페리아두'는 원하시는 컬러, 사이즈, 원단, 구성 등 다양하게 선택 가능합니다.

### 10.5.4. 주의 사항

- 명확히 인원수를 나타내는 경우에만 태깅한다.
- {n+[사람]에 속하는 보통명사}는 QT\_MAN\_COUNT로 태깅한다.
  - 편의점 앞에서 두 남자가 싸우고 있었다. → 편의점 앞에서 <두 남자:

QT\_MAN\_COUNT>가 싸우고 있었다.

- 한 꼬마, 두 꼬마, 세 꼬마 인디언 → <한 꼬마: QT\_MAN\_COUNT>, <두 꼬마: QT\_MAN\_COUNT>, <세 꼬마: QT\_MAN\_COUNT> 인디언
- 세 타자 연속 삼진 → <세 타자: QT\_MAN\_COUNT> 연속 삼진
- 한일 두 정상 간의 만남에 세계의 이목이 집중되고 있다. → 한일 <두 정상: QT\_MAN\_COUNT> 간의 만남에 세계의 이목이 집중되고 있다.

## 10.6. QT\_WEIGHT

### 10.6.1. 정의 및 설명

- 사물의 무거운 정도를 나타내는 표현
  - 무게

### 10.6.2. 분석 대상

- 무게
  - 100밀리그램(mg), 1,000그램(g), 1킬로그램(kg), 0.1톤(t), 1e-6킬로톤(kt), 584그레인(gr), 62온스(oz), 23파운드(lb), 7돈, 7냥, 1근, 7관, 30 kg, 140g, 백 근, 일만 톤, 170~300t , 한 돈

### 10.6.3. 예문

- 무게
  - 키 1m 80cm의 노모스카는 몸무게가 고작 <49kg: QT\_WEIGHT>이다.
  - 여기서 가장 큰 타이어인데 지름이 4m이고 무게가 <5 톤: QT\_WEIGHT>이래요.

### 10.6.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 2온스(oz) → <2온스: QT\_WEIGHT>(oz)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 100냥(3,750g) → <100냥: QT\_WEIGHT>(<3,750g: QT\_WEIGHT>)

## 10.7. QT\_PERCENTAGE

### 10.7.1. 정의 및 설명

- 비율과 관련된 표현



- 백분율, 비율
- 경쟁률
- 농도

### 10.7.2. 분석 대상

- 백분율, 비율
  - 75%, 80~90%, 1.1%, 0.719 %, 1/4, 1/100, 4분의 1, 4배, 2.5배, 열 배, 18.5:9. 4할 등
  - (연비) 1km/l , 2.352146mi/g, 100 L/100km 등
- 경쟁률
  - 3천 대 1, 50:1 등
- 농도
  - 50%, 56mg/L, 8000g/L, 2ppm, 10ppb, 400~500ppm 등
  - (주류 등의 알콜 도수) 40도, 12%, 196 proof, 151° 등

### 10.7.3. 예문

- 백분율, 비율
  - 자동차 타이어에는 <60% 정도: QT\_PERCENTAGE>의 합성고무가 들어가고 트럭 타이어에는 이보다 약간 덜 들어간대요.
  - 보통 1년에 공기 압력이 <4분의 1 정도: QT\_PERCENTAGE>가 줄어듭니다.
  - 디스플레이 비율이 <18.5:9: QT\_PERCENTAGE>인 아몰레드
  - 50L로 500km를 주행한다고 치면 연비는 <10km/L: QT\_PERCENTAGE>입니다.
- 경쟁률
 

<3천 대 1: QT\_PERCENTAGE>의 경쟁률을 뚫고 입사하였다.
- 농도
  - 당분의 농도가 <50% 이상: QT\_PERCENTAGE>이 되면 미생물의 발육이 억제된다.
  - 처음처럼의 도수는 <16.9도: QT\_PERCENTAGE>입니다.
  - 테라 맥주는 <4.6%: QT\_PERCENTAGE>라고 병뚜껑에 친절하게 적혀있더라고요. 다른 맥주들도 보통 <4~5%: QT\_PERCENTAGE> 정도 하지 않나요?
  - 바카디 151은 우리나라 수입 주류 중에서 제일 썸 술로서 도수가 무려 <151°: QT\_PERCENTAGE>(<75.5도: QT\_PERCENTAGE>)이다.

### 10.7.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 40퍼센트(%) → <40퍼센트: QT\_PERCENTAGE>(<40퍼센트: QT\_PERCENTAGE>)(%)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 10km/L(23.521458mi/g) → <10km/L: QT\_PERCENTAGE>(<23.521458mi/g: QT\_PERCENTAGE>)

## 10.8. QT\_SPEED

### 10.8.1. 정의 및 설명

- 속도 관련 표현

### 10.8.2. 분석 대상

- 속도
  - 1km/s, 30km/s, 1m/s, 3600m/h, 0.001km/s, 3.6km/h, 39.370079in/s, 141732.283in/h, 3.28084ft/s, 11811.0236ft/h, 0.000621mi/s, 2.236936mi/h, 1.943844노트(kn), 0.002941마하(mach)

### 10.8.3. 예문

- 속도
  - <시속 100 km까지: QT\_SPEED> 달리다가 브레이크를 밟는 거군요.
  - 이들 중성자는 믿기지 않을 정도의 엄청난 속도로 무려 <초속 2만 킬로미터: QT\_SPEED>로 피복을 뚫고 나오면 물 때문에 속도가 느려지면서 옆의 연료봉으로 뚫고 들어가죠.
  - 15초 동안 147m를 갔을 때 속도는 <9.8m/s: QT\_SPEED>입니다.

### 10.8.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 1마하(mach) → <1마하: QT\_SPEED>(mach)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 660.907127노트(1마하) → <660.907127노트: QT\_SPEED>(<1마하: QT\_SPEED>)

## 10.9. QT\_TEMPERATURE

### 10.9.1. 정의 및 설명

- 뜨거움과 차가움의 정도
  - 온도, 기온, 수온

### 10.9.2. 분석 대상

- 온도, 기온, 수온

- 22~25°C, -2.0°C, 9°C~12°C, 41~43.5도, 1섭씨온도(°C), 33.8화씨온도(°F), 274.15 절대온도(K), 493.47°R

### 10.9.3. 예문

- 온도, 기온, 수온
  - 핵분열로 뜨거워진 물은 원자로 내부에서 <320도까지: QT\_TEMPERATURE> 올라갑니다.
  - <800°C부터: QT\_TEMPERATURE>는 연료를 감싸고 있는 피복에 금이 가면서 방사선 물질이 갈라진 틈 사이로 빠져 나갑니다.
  - 화씨 <100°F: QT\_TEMPERATURE>는 섭씨 <37.8°C: QT\_TEMPERATURE>로 인간의 체온과 비슷하다.
  - 수온은 <80°C: QT\_TEMPERATURE>로 우리나라에서 수온이 가장 높은 온천의 하나이며, 대표적인 라듐 온천이다.

### 10.9.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 800도(°C) → <800도: QT\_TEMPERATURE>(°C)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 1°C(33.8°F) → <1°C: QT\_TEMPERATURE>(<33.8°F: QT\_TEMPERATURE>)

## 10.10. QT\_VOLUME

### 10.10.1. 정의 및 설명

- 물체나 물질이 공간에서 차지하는 크기
  - 부피, 용량, 용적

### 10.10.2. 분석 대상

- 부피, 용량, 용적
  - 2 만 2000km³, 20ℓ, 2,000~2,500mm, 0.15μg/ℓ, 1000시시(cc), 1000밀리리터(mℓ), 10데시리터(dℓ), 1리터(ℓ), 1000세제곱센티미터(cm³), 0.001세제곱미터(m³), 61.023744세제곱인치(in³), 0.035315세제곱피트(ft³), 0.001308세제곱야드(yd³), 0.264172갤런(gal), 0.006293배럴(bbl), 33.814022온스(oz), 5.543545홉, 0.554354되, 0.055435말, 두 말
  - (강수량) 100mm, 1000~2000mm

### 10.10.3. 예문

- 부피, 용량, 용적

- 올리브유 <1리터: QT\_VOLUME>를 얻으려면 적어도 올리브 5kg은 해야 해요.
- 이곳에서는 <1세제곱미터: QT\_VOLUME> 당 먼지 입자를 33개 미만으로 유지해야 하는데 <1세제곱미터: QT\_VOLUME>는 보통 냉장고만 한 공간이에요.
- 'fl.oz.'는 'Fluid Ounce'의 약자이고, 액체 부피를 나타내는 미국/영국식 파운드 개념의 측정방법이다.
- <1 fl.oz: QT\_VOLUME>는 <약 29.57 ml: QT\_VOLUME>이다.
- 연강우량의 최고는 인도 아셈 지방의 <10,000mm: QT\_VOLUME>이고, 우리나라의 총 강우량은 연 <600~1,500mm: QT\_VOLUME>이다.

#### 10.10.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 1갤런(gal) → <1갤런: QT\_VOLUME>(gal)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 1말(100홉) → <1말: QT\_VOLUME>(<100홉: QT\_VOLUME>)
- (강수량, 강우량) 강수량/강우량은 QT\_LENGTH의 'mm'와 구분하여 QT\_VOLUME으로 태깅한다.
  - 내일 서울에는 시간당 100mm의 비가 내리겠습니다. → 내일 서울에는 시간당 <100mm: QT\_VOLUME>의 비가 내리겠습니다.
  - 펫모드의 항공컨넬은 가로 350mm, 세로 500mm, 높이 650mm입니다. → 펫모드의 항공컨넬은 가로 <350mm: QT\_LENGTH>, 세로 <500mm: QT\_LENGTH>, 높이 <650mm: QT\_LENGTH >입니다.

### 10.11. QT\_ORDER

#### 10.11.1. 정의 및 설명

- 순서/순차적 표현
  - 순서, 순차
  - 등급

#### 10.11.2. 분석 대상

- 순서, 순차
  - 서수: 첫 번째, 두 번째, 세 번째, 일차, 이차, 삼차, 제7차, 1번, 2번, 십일 번 등
  - n학년/학기: 1~6학년, 초1, 초6, 중1, 중2, 고1, 고삼 / 1학기, 2학기, 3학기, 4학기 등
  - n주년/주기(周忌): 일 주년, 100주년, 백주년/일 주기, 100주기, 백주기 등
  - n기: 1기, 삼십오 기 등

- n호/권: 제82호, 80호/ 제 8권, 10권/vol 2 등
- n회/회차: 제10회/10회차 등
- 기타: 제4교구, 제2대, 12대, 제1대대, 일층, 3 층, 5공구, 제7광구, 삼수, 4수, 4옥타브, 1 코스, 2코스, 1구역, 2구역, 형사1부, 민사합의42부, 1호, 二號 등
- D+n: D+234, D+6 등
- 등급
  - n위/등: 1위, 2위 등
  - n레벨/등급/단계: 90,000 레벨, LV 1/ 1등급, 이 등급, 1단계, 십 단계 등
  - 한우 등급: 1, 2, 3 등급, 1+, 1++, 1++A, 1++B, 1++C, 2A, 2B, 2C, 3A, 3B, 3C, 투별 (투별) 등

### 10.11.3. 예문

- 순서, 순차
  - 이날 방송에서는 '여은파' <첫 번째: QT\_ORDER> 프로젝트 '홈트 프로젝트' 포스터 촬영 현장이 공개되었다.
  - 한국 학년제의 경우 초등학교는 <1학년~6학년: QT\_ORDER>, 중학교·고등학교는 <1학년~3학년까지: QT\_ORDER> 센다. 학점제인 고등교육기관에서도 학년을 세는데, 전문대학은 <1~4학년: QT\_ORDER>, 대학은 <1~5학년까지: QT\_ORDER> 세기도 하며, 대학원은 석사나 박사 과정에 따라 달라진다.
  - <고3: QT\_ORDER>이 지금 게임을 할 때야!
  - 국립부산과학관(관장 고현숙)은 12월 개관 <3주년: QT\_ORDER>을 맞아 다양한 이벤트, 전시, 공연 등 풍성한 프로그램을 마련하고 신규 전시물도 선보인다.
  - 3·1운동, 임정 <百주년: QT\_ORDER> 기념식이 서대문형무소에서 개최되었다.
  - 내일이 할아버지의 <이십오 주가: QT\_ORDER>가 되는 날이다.
  - [하루 한 권 그림책] <17기: QT\_ORDER> 모집 / 이번 기수 함께 읽을 엄마 책도 소개합니다.
  - 이 책의 체재 내용을 보면, <권1~2: QT\_ORDER>는 제목이 없고, <권3~5: QT\_ORDER>는 근포집(芹曝集), <권6~14: QT\_ORDER>는 진사록(辰巳錄), <권15~16: QT\_ORDER>은 군문등록(軍門騰錄), 난후잡록(亂後雜錄) 등으로 되어 있다.
  - 서울시민회관의 화재로 인해 <제10회: QT\_ORDER> 시상식은 국립극장에서 열렸다.
  - 스트라이퍼의 'In god we trust' 2017년 버전 커버 영상에서 <4옥타브: QT\_ORDER> 레 # (D#6) 음을 내고, 콘서트 라이브로도 이게 가능하다.
  - 서울고법 형사<1부: QT\_ORDER>(부장 정준영)는 '회복적 사법' 개념을 적용한 판결로 사회에 파장을 일으켜 왔다.
  - 초등학생 1학년 <D+6: QT\_ORDER>, 오늘은 남산 나들이를 했습니다.
- 등급

- <7위: QT\_ORDER>는 무한도전듀엣가요제.
- 거대한 힘의 반지, <9만 레벨: QT\_ORDER>을 찍었습니다!
- 그런데 제이미, 이게 아직 <1단계: QT\_ORDER>에 불과하다네요.
- 고깃집의 주메뉴는 한우 <투뿔: QT\_ORDER>입니다.
- 1% 한우 꽃등심 (180g) 42,000원, <1++: QT\_ORDER> 한우 차돌 26,000원, 생 갈비살 (180g) 20,000원!

#### 10.11.4. 주의 사항

- (원자번호) 원자번호는 QT\_ORDER로 태깅한다.
  - 원자 번호란 주기율표(periodic table)에서 각 원소마다 주어진 원소 고유의 순번으로, 원자 번호 1번 수소(Hydrogen, H)부터 118번 오가네손(Oganesson, Og)까지 있다. → 원자 번호 <1번: QT\_ORDER> 수소(Hydrogen, H)부터 <118번: QT\_ORDER> 오가네손(Oganesson, Og)까지 있다.
- (인쇄출판) 1쇄, 2쇄, 1권, 2판 등은 QT로 태깅한다. 단, 증보판, 재판은 태깅하지 않는다.
  - 한국어 문법론의 개념어 연구 초판 1쇄 발행 2012년 11월 23일 → 한국어 문법론의 개념어 연구 초판 <1쇄: QT\_ORDER> 발행 2012년 11월 23일
- 첫은 관형사이므로 단독으로는 태깅하지 않으며, 첫+보통명사도 태깅하지 않는다. (사전 등재 여부와 관련 없이 일괄적으로 태깅하지 않는다.)
  - 첫 여행, 첫 출발, 첫차 등
- 첫+날짜 관련 보통명사는 DT로 태깅한다.
  - 첫날, 첫 주말 → <첫날: DT\_DAY>, <첫 주말: DT\_DURATION>
- (특허분야) 특허분야는 이전의 도면과 청구항을 나타내는 패턴이 존재하며, 특정한 기능, 장치 등의 뒤에 괄호로 번호가 명시되는 경우가 존재한다.
  - 도1, 도 2, 청구항 1, 청구항 2 → <도1: QT\_ORDER>, <도 2: QT\_ORDER>, <청구항 1: QT\_ORDER>, <청구항 2: QT\_ORDER>
  - ~장치(100), ~모듈(200) → ~장치(<100: QT\_ORDER>), ~모듈(<200: QT\_ORDER>)
- (게임레벨) 레벨과 관련하여 '만렙'은 주의해서 태깅한다. 숫자로 쓴 LV10,000, 10,000 레벨, 1만 렙 등은 QT\_ORDER로 태깅, 한글로 쓴 만렙(滿LV)은 레벨 최대치를 채웠다는 의미이므로 태깅하지 않는다.
  - 스켈레톤으로 1만 렙 클리어! 초월 성공! → 스켈레톤으로 <1만 렙: QT\_ORDER> 클리어! 초월 성공!
  - 드디어 만렙 찍었어요!
- (다의적 쓰임) 맥락에 맞게 태깅한다.
- 회차를 의미하는 n회는 QT\_ORDER로 태깅하지만, 회수를 의미하는 n회는 QT\_COUNT로 태깅한다.
  - 창기배 10회는 2013년 10월 22일 ~ 10월 24일에 결승전이 진행되었고 스웨 9단이

- 우승을 차지했다. → 창기배 <10회: QT\_ORDER>는 2013년 10월 22일 ~ 10월 24일에 결승전이 진행되었고 스웨 9단이 우승을 차지했다.
- 日 주간지 “아베, 약효 떨어져 10회 시술中... 후계는 스가” → 日 주간지 “아베, 약효 떨어져 <10회: QT\_COUNT> 시술中... 후계는 스가”
  - ‘-째’가 ‘순서, 차례’의 의미일 때만 QT\_ORDER, ‘동안’의 의미일 경우에는 DT로 태깅한다.
    - 3일째 비가 오고 있었다. → <3일째: DT\_DURATION> 비가 오고 있었다.
    - 3일째 되던 날, 비가 그쳤다. → <3일째: QT\_ORDER> 되던 날, 비가 그쳤다.
  - 책의 권수는 QT\_COUNT로 태깅한다.
    - 징비록은 현재 4종이 전하는데 저자 자신의 필사원본인 『초본징비록(初本懲毖錄)』과 16권으로 된 『징비록』, 2권으로 된 간본(刊本), 필사본이 있다. → 징비록은 현재 4종이 전하는데 저자 자신의 필사원본인 『초본징비록(初本懲毖錄)』과 <16권: QT\_COUNT>으로 된 『징비록』, <2권: QT\_COUNT>으로 된 간본(刊本), 필사본이 있다.
  - [n일차/n년차] 日次가 ‘날의 차례’를 의미하므로 DT가 아닌 QT로 태깅한다.
    - 교생실습 1일차 → 교생실습 <1일차: QT\_ORDER>
    - 입사 8년차 → 입사 <8년차: QT\_ORDER>
  - [과목명+n] 과목명은 FD로, ‘n’은 QT로 태깅한다. (‘생1, 생2’와 같이 줄임말로 나와도 ‘생물’ 과목임을 알 수 있는 맥락이면 <생: FD\_SCIENCE>과 같이 줄임말 그대로 태깅한다.)
    - 생물1, 생물2 → <생물: FD\_SCIENCE><1: QT\_ORDER>, <생물: FD\_SCIENCE><2: QT\_ORDER>
    - 생1과 생2의 난이도 차이가 상당히 컸다. → <생: FD\_SCIENCE><1: QT\_ORDER>과 <생: FD\_SCIENCE><2: QT\_ORDER>의 난이도 차이가 상당히 컸다.
  - (목차) 장, 절, 항 등에 대해서는 각각 태깅하는 것을 기본으로 하되, 아래와 같은 경우에는 묶어서 태깅한다.
    - 시 69:5을 소리내어 읽어 봅시다. → 시 <69:5: QT\_ORDER>을 소리내어 읽어 봅시다.
    - 시편 제 69편 5절 → 시편 <제 69편: QT\_ORDER> <5절: QT\_ORDER>
  - [D+n(일)] 구성의 경우에는 QT\_ORDER로 태깅하는 것을 원칙으로 하고, ‘생후 n일’의 의미를 가질 때에 한해 QT\_AGE로 태깅한다.
    - 초등학생 1학년 D+6, 오늘은 남산 나들이를 했습니다. → 초등학생 1학년 <D+6: QT\_ORDER>, 오늘은 남산 나들이를 했습니다.(초등학생이 된 지 여섯 번째 날에 남산 나들이를 했다고 해석)
    - 빈나네 마당냥이 4남매 D+100!! 백일잔치~츄르잔치~ → 빈나네 마당냥이 4남매 <D+100: QT\_AGE> 백일잔치~츄르잔치~ (문맥상 생후 n일째의 의미이므로 나이로 해석)
  - (학년/반)
    - n학년 n반: 학년과 반을 따로 태깅한다.

- 서천초등학교 6학년 1반 친구들의 사랑이 넘치는 가족공예체험수업! → 서천초등학교 <6학년: QT\_ORDER> <1반: QT\_ORDER> 친구들의 사랑이 넘치는 가족공예체험 수업!
- 단, '6-1'과 같이 [n-n]의 구성은 전체를 묶어 QT\_ORDER로 태깅한다.
  - 현대고등학교 3-1 정미소의 가방. → 현대고등학교 <3-1: QT\_ORDER> 정미소의 가방.

## 10.12. QT\_PRICE

### 10.12.1. 정의 및 설명

- 어떤 대상에 일정하게 매겨진 액수
  - 값, 가격, 금액

### 10.12.2. 분석 대상

- 값, 가격, 금액
  - 72만 원, 60억 달러, \$ 40,000, 403원, 290원, 1조 2천억 원, ₩188,000, 100만 원, 백만 원

### 10.12.3. 예문

- 값, 가격, 금액
  - 소더비 경매시장에서 무려 <200만 달러: QT\_PRICE>, <약 26억원: QT\_PRICE> 되는 돈에 낙찰됐습니다.
  - 윤상현 씨가 전소속사로부터 <10억 원대: QT\_PRICE>의 손해배상소송에 휩싸였습니다.

### 10.12.4. 주의 사항

- 단위가 붙어있지 않아도 맥락상 금액을 의미하면 QT\_PRICE로 태깅한다.
  - 30평도 안 되는데 17억이 말이 되냐? → 30평도 안 되는데 <17억: QT\_PRICE>이 말이 되냐?
  - 한 시간 통역하면 15~20정도 받아. → 한 시간 통역하면 <15~ 20정도: QT\_PRICE> 받아.

## 10.13. QT\_PHONE

### 10.13.1. 정의 및 설명



- 전화번호
  - 휴대전화번호
  - 일반전화번호
  - 긴급상담전화번호 등 공공기관번호
  - 기타-내선번호, 국번, 국제전화 국가번호 등

### 10.13.2. 분석 대상

- 휴대전화번호
  - 010-1111-1234, 001821012345678
- 일반전화번호
  - 02)123-1234, 032-3456-7890, 0070082234561234
- 긴급상담전화번호
  - 120, 114, 112, 119, 1366
- 기타
  - 101번, 9번, #1230
  - 1588-1588
  - 031, 02, 001, 00700, 080, 81, 82

### 10.13.3. 예문

- 휴대전화번호
  - 제 핸드폰 번호는 <010-1111-1234: QT\_PHONE>입니다.
  - 국제전화 서비스번호 아무거나 누르고 <821012345678: QT\_PHONE> 눌러.
- 일반전화번호
  - 이들 행사의 참가비는 없으며, 선착순 입장해야 한다. (문의: ☎<02) 345-6313: QT\_PHONE>)
- 긴급상담전화번호
  - 긴급통신(<119: QT\_PHONE>, <112: QT\_PHONE>) 및 감청방식이 가닥이 잡혔다.
  - 여성긴급전화 <1366: QT\_PHONE>
- 기타
  - 기획조정실의 내선번호는 <101번: QT\_PHONE>, 비서실은 <102번: QT\_PHONE>입니다.
  - 앞뒤가 똑같은 전화번호, <1577-1577: QT\_PHONE>
  - 일본 국가번호가 <81: QT\_PHONE>이고 우리나라는 <82: QT\_PHONE>라고.
  - 경기도가 <031: QT\_PHONE>이야, <032: QT\_PHONE>야?
  - 아직도 <011: QT\_PHONE> 쓰는 사람이 있어?

## 10.14. QT\_SPORTS

### 10.14.1. 정의 및 설명

- 스포츠 관련 수량 표현

### 10.14.2. 분석 대상

- 스포츠 관련 수량 표현
  - 1승2패, 5타점, 3:2, 2연승, 한 판, 두 판
  - ~경기, ~승/패/무, 10 호골, ~안타

### 10.14.3. 예문

- 스포츠 관련 수량 표현
  - 아스날은 이번 경기에서의 승리로 칼링컵 포함 <14경기: QT\_SPORTS>에서 <13승 1무: QT\_SPORTS>를 하며 무패 행진을 이어 가게 되었다.
  - 아데바요르는 왼쪽으로 가볍게 성공시키며 <1:0: QT\_SPORTS>으로 앞서게 되었고 그 골을 그대로 지켜내며 결승골로 만들어 <1:0: QT\_SPORTS> 승리를 거두었다.
  - 임태혁은 첫 <두 판: QT\_SPORTS>을 먼저 따내 승기를 잡는 듯했습니다.

### 10.14.4. 주의 사항

- (스포츠규칙/용어) 볼넷, 삼진과 같은 스포츠 규칙 관련 용어나 수량의 의미가 없는 1, 2루와 1, 2레인은 TM\_SPORTS로 태깅한다.
  - 득점 기회였는데 NC 박석민이 삼진으로 물러나면서 이닝이 종료되었다. → 득점 기회였는데 NC 박석민이 <삼진: TM\_SPORTS>으로 물러나면서 이닝이 종료되었다.
- (스포츠규칙/용어) 6이닝, 1볼넷, 6오버파의 경우 QT로 태깅한다.
  - 7월 21일 경기에서 SK 김광현이 6이닝 2실점으로 승리 투수가 되었습니다. → 7월 21일 경기에서 SK 김광현이 <6이닝: QT\_SPORTS> <2실점: QT\_SPORTS>으로 승리 투수가 되었습니다.
- (스포츠규칙/용어) '한 판'과 '한판(유도의 판정 용어)'은 각각 QT와 TM으로 태깅한다. 띄어쓰기 참조.
  - 한 판만 더 이기면 강호동이 우승이다. → <한 판만: QT\_SPORTS> 더 이기면 강호동이 우승이다.
  - 엮어치기 할 때 중심 다리가 상대방 다리 사이로 들어가면 절대 한판이 안 나온다. → 엮어치기 할 때 중심 다리가 상대방 다리 사이로 들어가면 절대 <한판: TM\_SPORTS>이 안 나온다.
- 'n관왕'은 특정 스포츠 관련 용어가 아니고 '아카데미 3관왕'과 같이 다른 분야에서도 사용할 수 있기 때문에 QT\_COUNT로 태깅한다.
  - 수원시청은 김나리-홍승연이 복식에도 결승에 올라, 단체전 우승에 이어 3관왕을 바라보

게 됐다. → 수원시청은 김나리-홍승연이 복식에도 결승에 올라, 단체전 우승에 이어<3관왕: QT\_COUNT>을 바라보게 됐다.

## 10.15. QT\_CHANNEL

### 10.15.1. 정의 및 설명

- 텔레비전 · 라디오 · 무선 통신 따위의 주파수, 채널 번호
  - 공중파/케이블 TV
  - 라디오/무선 통신 등

### 10.15.2. 분석 대상

- 공중파/케이블 TV
  - 채널 6, 44번 채널, 0번, 200번
- 라디오/무선 통신 등
  - 91.9, 95.9 MHz

### 10.15.3. 예문

- 공중파/케이블 TV
  - tvN의 채널번호가<17번: QT\_CHANNEL>에서 <3번: QT\_CHANNEL>으로 변경됩니다.
- 라디오/무선 통신 등
  - MBC라디오 FM4U의 주파수는 <91.9MHz: QT\_CHANNEL>입니다.

### 10.15.4. 주의 사항

- 명확한 채널번호에 해당되는 경우에만 태깅한다.
- 39 쇼핑은 케이블 채널 39번에서 방영됩니다. → 39 쇼핑은 케이블 채널 <39번: QT\_CHANNEL>에서 방영됩니다. (이때 '39 쇼핑'은 채널명이므로 AFA\_VIDEO로 태깅)
- (프로그램명) CJ홈쇼핑, KBS 2TV의 경우는 채널명에 해당되기 때문에 AFA\_VIDEO(프로그램명)으로 태깅한다.
  - CJ홈쇼핑, 코미디 TV, MBCsports+ → <CJ홈쇼핑: AFA\_VIDEO>, <코미디 TV: AFA\_VIDEO>, <MBCsports+: AFA\_VIDEO>
- (방송기관) 문맥에서 채널명이 아닌 방송기관으로 판단될 경우 기관명에 해당하는 OG로 태깅한다.
  - 39쇼핑은 CJ오쇼핑에 인수되었습니다. → <39쇼핑: OGG\_ECONOMY>은 <CJ오쇼핑: OGG\_ECONOMY>에 인수되었습니다.

## 10.16. QT\_ALBUM

### 10.16.1. 정의 및 설명

- 노래 또는 연주곡 따위의 음반과 관련된 순차 표현
  - 테이프, 레코드판, 시디, 음원파일 등

### 10.16.2. 분석 대상

- 테이프, 레코드판, 시디, 음원파일 등
  - 1집, 1.5집, VI, part 1, vol. 2 등

### 10.16.3. 예문

- 테이프, 레코드판, 시디, 음원파일 등
  - <5집: QT\_ALBUM> 발매 보아, '남성 우월주의 거부한다'
  - 작곡가 최희주와 작사가 고은주가 주축이 돼 이끌고 있는 프로젝트 그룹 DnD가 여름을 맞아 싱글 앨범 <2.5집: QT\_ALBUM>을 발매했다.
  - YUN SANG <PART1: QT\_ALBUM>은 상업적으로나 대중적 인기 면에서 두루 성공한 윤상의 최고 히트작이었다. 그러나 <PART2: QT\_ALBUM>는 다양한 음악적 시도에 초점을 맞췄기에 상업성과 거리가 멀어졌다.

### 10.16.4. 주의 사항

- (음반) 앨범명은 AFW\_OTHER\_PRODUCTS로 태깅하고 'N집'을 의미하는 경우에는 QT\_ALBUM으로 태깅한다.
  - 푸른하늘의 3집 푸른하늘의 수록곡 푸른하늘 → <푸른하늘: PS\_NAME>의 <3집: QT\_ALBUM> <푸른하늘: AFW\_OTHER\_PRODUCTS>의 수록곡 <푸른하늘: AFA\_MUSIC>

## 10.17. QT\_ADDRESS

### 10.17.1. 정의 및 설명

- 주소와 관련된 숫자 표현
  - 사서함 번호
  - 우편번호
  - 주소의 상세 번지 및 공동주택의 개별 호수 등

### 10.17.2. 분석 대상

- 사서함 번호
  - 76-8호, 20호, GZ00000000KR, KR0178655117
- 우편번호
  - 110-704, 04766, 97251-9610
- 주소의 상세 번지 및 공동주택의 개별 호수 등
  - 5, 9-30, 668-4, 1102호, 702호, 179번지

### 10.17.3. 예문

- 사서함 번호
  - 논산훈련소 23연대 주소: 충남 논산시 연무읍 사서함 <76-8호: QT\_ADDRESS>
- 우편번호
  - (<760-905: QT\_ADDRESS>) 경북 안동시 태화동 677-6
  - (<04766: QT\_ADDRESS>) 서울특별시 성동구 서울숲9길 9-7 (성수동1가)
- 주소의 상세 번지 및 공동주택의 개별 호수 등
  - 서울 동대문구 답십리로 <130: QT\_ADDRESS> 답십리래미안위브 401동 <701호: QT\_ADDRESS>

### 10.17.4. 주의 사항

- (예외) 도로명에 붙은 'n길', 'n로' 등은 도로명과 묶어서 AF\_ROAD로, 아파트명 등에 붙은 'n단지', 'n차'는 아파트명, 'n동'과 함께 묶어 AF\_BUILDING으로 태깅한다. 단, 아파트명 등에 'n단지', 'n차'만 붙어 쓰였을 경우에는 LC\_OTHERS로 태깅한다.
  - 서울숲9길 9-7 → <서울숲9길: AF\_ROAD> 9-7
  - 테헤란로 7길 8 → <테헤란로 7길: AF\_ROAD> 8
  - 잠실주공아파트 5단지 524동 802호 → <잠실주공아파트 5단지 524동: AF\_BUILDING> 802호
  - 장미 2차 아파트 29동 505호 → <장미 2차 아파트 29동: AF\_BUILDING> 505호
  - 래미안서초 6차에 전세 나온 거 있어요? → <래미안서초 6차: LC\_OTHERS>에 전세 나온 거 있어요?
- 'n-n' 형식의 상세 주소는 하나로 태깅하고 'n번지', 'n호' 등의 상세 주소는 각각 따로 태깅한다.
  - 강남구 언주로 551-1 → 강남구 언주로 <551-1: QT\_ADDRESS>
  - 강남구 대치동 한보은마아파트 101동 1102호 → 강남구 대치동 한보은마아파트 101동 <1101호: QT\_ADDRESS>
- (주소 태깅의 예)
  - (06288) 서울시 강남구 삼성로 1길 150 대치미도 1차 아파트 103동 101호 → (<06288: QT\_ADDRESS>) <서울시: LCP\_CAPITALCITY> <강남구: LCP\_COUNTY> <삼성로 1길:

AF\_ROAD> <150: QT\_ADDRESS> <대치미도 1차 아파트 103동: AF\_BUILDING> <101호: QT\_ADDRESS>

- 단, 공동주택의 동/호수를 'n-n'의 형식으로 묶어 쓴 경우에는 아래와 같이 QT\_ADDRESS로 태깅한다.
- 서울시 강남구 삼성로 1길 150 대치미도 1차 아파트 103-101 → 서울시 강남구 삼성로 1길 150 <대치미도 1차 아파트: LC\_OTHERS> <103-101: QT\_ADDRESS>

## 10.18. QT\_OTHERS

### 10.18.1. 정의 및 설명

- 기타 수량 표현
  - 10.1~10.17까지의 세부 분류에 해당하지 않는 기타 수량 표현

### 10.18.2. 분석 대상

- 기타 수량
  - 점수: 45점, 100점, 87/100(100점 만점에 87점의 의미)
  - 백분위: 백분위등급 75, 백분위 80, 백분위 97
  - 경도/위도: 북위 38 53' 51, 동경 131도 52분 8.3406초
  - 버스 번호: 20번, 1112번, M5107
  - 열차 번호: 165, 452, 706
  - 해상도: 2960X1440
  - 데이터양: 8388608비트(bit), 1048576바이트(B), 1024킬로바이트(KB), 1메가바이트(MB), 0.000977기가바이트(GB), 9.5367e-7테라바이트(TB), 9.3132e-10페타바이트(PB), 9.0949e-13엑사바이트(EB)
  - 압력: 1기압(atm), 101325파스칼(Pa), 1013.25헥토파스칼(hPa), 101.325킬로파스칼(kPa), 0.101325메가파스칼(MPa), 1013250dyne/cm2, 1013.25밀리바(mb), 1.01325바(bar), 1.033227kgf/cm2, 14.696프사이(psi), 760수은주밀리미터(mmHg), 29.92126inchHg, 10332.275수주밀리미터(mmH2O), 406.782188inchH2O
  - 계좌번호: 12-345-6789-0
  - 자동차등의 번호판: 123가4567
  - 가상화폐: 1BTC, 10ETH, 100XRP, 1,000BCH, 10,000LCT
  - 기타: 지식머니, 게임머니(재화) 등

### 10.18.3. 예문

- 기타 수량

- 1등급 커트라인이 <45점: QT\_OTHERS>이었다.
- 어떤 학생의 원점수가 60점인데 백분위 등급이 <75: QT\_OTHERS>였다면 이 학생이 받은 60점 아래에 전체 학생의 75%가 있다는 의미가 되고 이 학생은 상위 25%에 해당한다.
- 독도의 위도와 경도는 <동경 131도 52분 8.3406초: QT\_OTHERS> <북위 37도 14분 22.2636초: QT\_OTHERS>이다.
- 잠실5단지 앞에서 <1112번: QT\_OTHERS>을 타면 됩니다.
- KTX나 SRT를 타다 보면 세 자리수 열차번호가 있을텐데 (예를 들어서 KTX 부산행 '<165: QT\_OTHERS>'열차) 이 세자리 수가 뭘 뜻하는지 궁금했을 겁니다.
- 모니터의 해상도를 <2960X1440: QT\_OTHERS>로 맞춰 주세요.
- <1기가바이트: QT\_OTHERS>는 <1,024MB: QT\_OTHERS>이다.
- 헥토파스칼(hPa)은 기상학에서 사용하는 기압의 단위를 말하며 <1hPa: QT\_OTHERS>은 <1Pa: QT\_OTHERS>의 100배이다.
- 계좌번호는 <101-04-3985-114: QT\_OTHERS>입니다.
- 비트코인은 <1BTC: QT\_OTHERS>(비트코인 단위)당 1,998만 9천 원에 거래돼 24시간 전보다 4.84% 하락하고 있다.
- 지식 머니는 얼마나 걸어야 할지는 모르겠지만 자세히 적어 주시면 <10 지식머니: QT\_OTHERS>를, 조금 어설프다 싶으면 <5 지식머니: QT\_OTHERS>를 드릴게요.
- 이벤트 기간 동안 '로만의 금빛 패키지1'을 구매하시면 기본 <500루비: QT\_OTHERS>에 <30,000,000골드: QT\_OTHERS>와 <11 루아의 증표: QT\_OTHERS>를 보너스로 드립니다.

#### 10.18.4. 주의 사항

- 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
  - 1기가바이트(GB) → <1기가바이트: QT\_OTHERS>(GB)
- 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
  - 1hPa(100Pa:QT\_OTHERS) → <1hPa: QT\_OTHERS>(<100Pa: QT\_OTHERS>)
- +, -, ÷, ×, / 의 수학적 기호가 쓰인 경우, 맥락에 상관 없이 QT\_OTHERS로 태깅한다.
  - 모카프라프치노 1+1 행사 중 → 모카프라프치노 <1+1: QT\_OTHERS>행사 중
  - 각대 봉투 20×50sets → 각대 봉투 <20×50sets: QT\_OTHERS>
  - 펫모드 스카이 항공켄넬 51×70×47 → 펫모드 스카이 항공켄넬 <51×70×47: QT\_OTHERS>
- 정체를 알 수 없는 불명확한 숫자는 태깅하지 않는다.
  - +254, #1#, \*\*\*12 등
- 노선버스의 번호에 붙는 영문은 태깅범위에 포함한다.
  - 동탄2신도시 M버스 M4434, G버스 G6009 신안인스빌리베라 1차 정차! → 동탄2신도시 M버스 <M4434: QT\_OTHERS>, G버스 <G6009: QT\_OTHERS> 신안인스빌리베라 1차 정

차!

## 11. EVENT

---

EVENT는 특정 사건/사고/행사 명칭을 나타낸다.

### 11.1. EV\_ACTIVITY

#### 11.1.1. 정의 및 설명

- 사회운동
- 선언

#### 11.1.2. 분석 대상

- 사회운동
  - 공민권운동, 브나로드운동, 새마을운동, 양무운동, 변법자강운동, 만세시위운동, 갑오개혁, 항일독립운동
- 선언
  - 인권선언, 노예해방선언

#### 11.1.3. 예문

- 사회운동
  - 1962년 흑인의 <투표권등록운동: EV\_ACTIVITY>이 시작되었다.
  - 1931년부터 1934년까지 동아일보사가 전개한 <브나로드운동: EV\_ACTIVITY>은 고등보통 학교 4, 5학년 학생으로 이루어진 학생계몽대와 전문학교 이상의 학생으로 조직된 학생 강연대, 학생기자대를 주축으로 하여 행해졌다.
  - <새마을운동: EV\_ACTIVITY>은 1970년대의 한국사회를 특징짓는 중요한 사건이다.
  - 그러나 이 <양무운동: EV\_ACTIVITY>이 중국 자본주의 발달의 싹이 되었다는 것은 부인할 수 없다.
- 선언
  - <인권선언: EV\_ACTIVITY>은 1891년의 프랑스 헌법을 비롯하여 그 후 세계 여러 나라의 헌법 및 정치에 크게 영향을 미쳤다.



#### 11.1.4. 주의 사항

- ~운동, ~선언 등이 해당된다.
- EV\_ACTIVITY은 시기, 기간이 명확한 구체적인 사건에 한정하여 태깅한다.
- (갑오개혁, 갑오경장) EV\_ACTIVITY에 해당된다.
  - 갑오개혁은 조선 고종 31년 7월부터 고종 33년 2월 사이에 추진되었던 개혁 운동이다.  
→ <갑오개혁: EV\_ACTIVITY>은 조선 고종 31년 7월부터 고종 33년 2월 사이에 추진되었던 개혁 운동이다.
  - 조선왕조는 갑오경장이라는 이름의 개혁을 통해 쓰러져 가는 국운을 되살리려 애썼다.  
→ 조선왕조는 <갑오경장: EV\_ACTIVITY>이라는 이름의 개혁을 통해 쓰러져가는 국운을 되살리려 애썼다.
- (독립운동) 독립운동에 해당하는 사건도 EV\_ACTIVITY에 해당된다.
  - 임시 정부도 항일 독립운동을 전개하기 위해서는 많은 자금이 필요했습니다. → 임시 정부도 <항일 독립운동: EV\_ACTIVITY>을 전개하기 위해서는 많은 자금이 필요했습니다.
- (~칙령) '~칙령'은 CV\_LAW에 해당된다.
  - 밀라노 칙령에는 그리스도교 공인에 관한 획기적인 명령이 들어있습니다. → <밀라노 칙령: CV\_LAW>에는 그리스도교 공인에 관한 획기적인 명령이 들어있습니다.
- (사보타주) 사보타주는 노동운동보다는 특정 행동으로 볼 수 있으므로 EV\_ACTIVITY의 대상이 아니다.
  - UPS의 사보타주를 보상하기 위해 무선 이어폰을 선물했다.

## 11.2. EV\_WAR\_REVOLUTION

### 11.2.1. 정의 및 설명

- 전쟁/혁명
- 전투

### 11.2.2. 분석 대상

- 전쟁/혁명
  - 러일전쟁, 남북전쟁, 포클랜드전쟁, 아편전쟁, 임진왜란, 청교도혁명, 4.19 혁명, 신해혁명, 메이지유신
- 전투
  - 명량해전, 칠천량해전, 노량해전, 연평해전, 봉오동 전투, 황산벌 전투, 워털루 전투

### 11.2.3. 예문

- 전쟁/혁명

- <러·일전쟁: EV\_WAR\_REVOLUTION>의 결과는 포츠머스 강화회담과 을사조약으로 이어져
- <남북전쟁: EV\_WAR\_REVOLUTION>은 19세기 미국에서 발생한 가장 주목할 만한 역사적 사건이었다.
- <4·19 혁명: EV\_WAR\_REVOLUTION>은 한국의 정치발전사에 하나의 획기적인 전기(轉機)를 기록한 역사적으로 큰 의미를 지니는 일대사건(一大事件)이었다.
- 전투
  - 당연히 <안시성 전투: EV\_WAR\_REVOLUTION>는 고구려-당나라 전쟁에서 최대의 클라이맥스가 됩니다.
  - 김흠춘의 아들인 반굴은 김유신과 부친과 함께 <황산벌 전투: EV\_WAR\_REVOLUTION>에 참전했다.

#### 11.2.4. 주의 사항

- ~전쟁/~혁명을 대상으로 한다.
- EV\_WAR\_REVOLUTION은 시기, 기간이 명확한 구체적인 사건에 한정하여 태깅한다.
- (전투작전) 전투작전도 EV\_WAR\_REVOLUTION에 포함된다.
  - 노르망디 상륙작전의 성공으로 연합군은 그해 8월 25일 파리를 탈환하였다. → <노르망디 상륙작전: EV\_WAR\_REVOLUTION>의 성공으로 연합군은 그해 8월 25일 파리를 탈환하였다.
- (전쟁명칭) '세계2차대전' 등의 전쟁 명칭도 포함된다.
- 세계2차대전은 20세기 전 인류에 최고의 영향을 미친 사건으로 기록될 것입니다. → <세계2차대전: EV\_WAR\_REVOLUTION>은 20세기 전 인류에 최고의 영향을 미친 사건으로 기록될 것입니다.
- (무력 정변) EV\_WAR\_REVOLUTION에 해당된다.
  - 1884년 12월 갑신정변 때, 민영익과 그 동료들이 큰 부상을 입게 된다. → 1884년 12월 <갑신정변: EV\_WAR\_REVOLUTION> 때, 민영익과 그 동료들이 큰 부상을 입게 된다.
  - 한국 현대사의 비극중의 하나인 5·16 쿠데타가 1961년 5월 16일 새벽에 발생했다. → 한국 현대사의 비극중의 하나인 <5·16 쿠데타: EV\_WAR\_REVOLUTION>가 1961년 5월 16일 새벽에 발생했다.
- (민란, 반란, 반정) EV\_WAR\_REVOLUTION에 해당된다.
  - 박규수는 1862년 임술민란 때는 백성을 무마하는 안핵사로 진주에 내려갔었다. → 박규수는 1862년 <임술민란: EV\_WAR\_REVOLUTION> 때는 백성을 무마하는 안핵사로 진주에 내려갔었다.
  - 청주에서 이인좌의 난이 일어난 것은 영조 4년(1728)이었다. → 청주에서 <이인좌의 난: EV\_WAR\_REVOLUTION>이 일어난 것은 영조 4년(1728)이었다.
  - 1623년 인조반정 성공 나흘 뒤 왕위에 오른 능양군 이종이 선언했다. → 1623년 <인조

- 반정: EV\_WAR\_REVOLUTION> 성공 나흘 뒤 왕위에 오른 능양군 이종이 선언했다.
- (농민봉기) EV\_WAR\_REVOLUTION에 해당된다.
    - 임술농민봉기는 단성에서 시작되고 곧 이어 진주로 파급되어갔다. → <임술농민봉기: EV\_WAR\_REVOLUTION>는 단성에서 시작되고 곧 이어 진주로 파급되어갔다.
  - (모반 사건) EV\_WAR\_REVOLUTION에 해당된다.
    - 정여립 모반 사건으로 인한 기축옥사로 동인이 대거 제거된다. → <정여립 모반 사건: EV\_WAR\_REVOLUTION>으로 인한 기축옥사로 동인이 대거 제거된다.
  - (4차산업혁명, 농업혁명) 시기, 기간을 특정할 수 없는 사조나 사회변화를 의미하므로 CV\_CULTURE로 태깅한다.
  - (프롤레타리아혁명) 시기를 특정할 수 없는, 사건이라기보다 개념에 해당하므로 태깅 대상이 아니다.

### 11.3. EV\_SPORTS

#### 11.3.1. 정의 및 설명

- 스포츠 관련 행사
- 게임 관련 e 스포츠 이벤트(행사) 이름
- 스포츠 관련 행사의 일반 유형

#### 11.3.2. 분석 대상

- 스포츠 관련 행사
  - ~대회/경기: 올림픽경기대회, 아시아경기대회, 서울 올림픽 경기, 88 서울올림픽, 서울올림픽, 거제시장기 전국 윈드서핑대회, 홈경기, 원정경기, 세계 역도 선수권 대회, 남아공 월드컵 축구대회, 내셔널리그 챔피언십시리즈, 평창 동계 올림픽, 바르셀로나 올림픽
  - ~전: 두산전, 개막전, 최종전, 결승전, 예선전, 전반기, 후반기, 조별예선전, 주중 3 연전, A매치, 호주전, 한국-호주전
- 게임 관련 e 스포츠 이벤트(행사) 이름
  - 오버워치 리그, 던전애파이터 프리미어 리그, LoL 월드챔피언십
- 스포츠 관련 행사의 일반 유형
  - 마라톤 대회, 등반 대회, 축구 대회, 자전거 대회, 테니스 대회, 태권도 대회, 조정 대회

#### 11.3.3. 예문

- 스포츠 관련 행사
- 지난 시즌에도 1라운드 4승 5패로 부진했지만 결국 <챔피언결정전: EV\_SPORTS>에 진출하는 뒷심을 발휘했다.

- 1989년 <마스터스: EV\_SPORTS> 연장 첫 번째 홀에서
- <LPGA 골프 대회: EV\_SPORTS>에서 우승하기란 얼마나 어려운 것인지.
- 특히 올해 열세 번째를 맞는 <수중 마라톤 대회: EV\_SPORTS>는 황강의 물줄기를 이용한 독특한 코스를 달리는 전국 유일의 마라톤 대회로 마라톤 마니아뿐 아니라 가족 단위의 참가자들에게도 인기를 끌고 있다.
- 지난해 <백두대간 그란폰도 대회: EV\_SPORTS>에 참가한 선수들 출발 모습.
- <전반전: EV\_SPORTS>, <후반전: EV\_SPORTS> 풀타임을 소화한 손흥민 선수
- 게임 관련 e 스포츠 이벤트(행사) 이름
  - 이번 생중계는 오는 21일 진행되는 <오버워치 리그: EV\_SPORTS> 3주차 경기를 시작으로 오는 9월 말까지 이어진다.
  - 중국 상하이에서 열린 '2020 <LoL 월드 챔피언십: EV\_SPORTS>' 결승전에서 담원 게이밍이 쑤닝을 꺾고 우승했다.
- 스포츠 관련 행사의 일반 유형
  - 해발 1천 567m 태백산 정상에 올라 눈 꽃을 감상하는 <등반대회: EV\_SPORTS>를 끝으로 막을 내리게 된다.
  - 유명 <마라톤 대회: EV\_SPORTS>는 참가신청 접수가 몇 분만에 마감되기도 한다.
  - 이번 저희 클럽에서 11월달 행사로 취미반 친구들을 위한 <축구대회: EV\_SPORTS>를 준비하였습니다.
  - 10월 말에 연천에서 DMZ를 통과하는 <자전거 대회: EV\_SPORTS>가 있다고 해서 가져와 봤어요!

#### 11.3.4. 주의 사항

- EV\_SPORTS는 일반 유형도 태깅한다.
- (라운드, 회, 세트, 경기, 강 ) 멀티 개체명 분류 가능하다. 스포츠 행사의 경우 '4강, 8강' 등과 같이 숫자가 나와도 문맥에 따라 QT가 아닌 EV로 처리해야 하는 경우가 있으므로 주의해야 한다.
  - (숫자+라운드) '2라운드'와 같은 표현이 경기의 수를 세는 의미라면 QT\_SPORTS로, 예선전, 결승전과 같이 경기 자체를 의미하는 것이라면 EV\_SPORTS로 태깅한다.
  - 이번 K리그 축구는 총 32라운드를 경기할 예정이다. → 이번 K리그 축구는 총 <32라운드: QT>를 경기할 예정이다.
  - KPGA 오픈 3라운드에서 김민규가 1위에 오르며 지난주 군산CC 오픈에서 우승한 김주형에 이어 10대 돌풍을 이어갈 기세다. → <KPGA 오픈 3라운드: EV\_SPORTS>에서 김민규가 1위에 오르며 지난주 군산CC 오픈에서 우승한 김주형에 이어 10대 돌풍을 이어갈 기세다.
  - 마지막 4라운드에는 비 소식이 예정되어 있다. → 마지막 <4라운드: EV\_SPORTS>에는 비 소식이 예정되어 있다.

- (숫자+회, 숫자+세트) 문맥에 준하여 태깅한다.
  - 한화는 7회까지 <2회: QT\_SPORTS>를 제외하고 매회 누상에 주자를 남겨둔 채 공격을 접어야 했다.
  - 추신수는 9회 마지막 타석에서 헛스윙 삼진으로 물러났으나 텍사스는 1점 차로 이겼다. → 추신수는 <9회: EV\_SPORTS> 마지막 타석에서 헛스윙 삼진으로 물러났으나 텍사스는 1점 차로 이겼다.
  - 대역전극은 9회에서 펼쳐졌다. → 대역전극은 <9회: EV\_SPORTS>에서 펼쳐졌다.
  - 돌아온 샤라포바, US오픈 3회전 진출 → 돌아온 샤라포바, <US오픈 3회전: EV\_SPORTS> 진출
  - 정현은 2회전에서 존 이스너에 0-3 패배 → 정현은 <2회전: EV\_SPORTS>에서 존 이스너에 0-3 패배
  - 역전 드라마는 마지막 5세트에서 펼쳐졌다. → 역전 드라마는 마지막 <5세트: EV\_SPORTS>에서 펼쳐졌다.
- (숫자+경기) 문맥에 준하여 태깅한다.
  - 이 사진은 2019 <세계 남자농구월드컵 조별예선 3경기: EV\_SPORTS>의 하이라이트 장면이다.
  - 아스날은 이번 경기에서의 승리로 칼링컵 포함 <14경기: QT\_SPORTS>에서 <13승 1무: QT\_SPORTS>를 하며 무패 행진을 이어 가게 되었다.
- (숫자+강) '4강, 8강' 등 앞에 숫자가 나온 경우, '4강전'과 같이 '~전'을 의미하면 EV\_SPORTS로, 4개의 강팀을 의미하면 QT\_SPORTS로 태깅한다.
  - 프로야구는 6강전, 5강전을 진행한다. → 프로야구는 <6강전: EV\_SPORTS>, <5강전: EV\_SPORTS>을 진행한다.
- (숫자+파전) EV\_SPORTS로 태깅한다.
  - 고양시 女축구대회, 한중일 3파전으로 열린다. → 고양시 女축구대회, 한중일 <3파전: EV\_SPORTS>으로 열린다.
- (개최 회수+지역+스포츠 행사) 스포츠 행사는 보통 개최 회수, 지역 등과 같이 쓰인다. 개최 회수는 분리하여 태깅하고 지역은 통합하여 태깅한다.
  - 제24회 서울 올림픽 경기 → <제24회: QT\_ORDER> <서울 올림픽 경기: EV\_SPORTS>
- (연도 표시+스포츠 행사) 스포츠 행사가 DATE와 병행하여 쓰일 경우 명확한 연도 표시는 독립하여 태깅하고, 축약된 경우는 통합하여 태깅한다.
  - 야마구치 이사는 1988년 서울 올림픽에 출전해 동메달을 획득했다. → 야마구치 이사는 <1988년: DT\_YEAR> <서울 올림픽 경기: EV\_SPORTS>에 출전해 동메달을 획득했다.
  - 88 서울올림픽경기 16일간의 대장정을 알리는 성화가 막 점화되어 타오르고 있다. → <88 서울올림픽경기: EV\_SPORTS> 16일간의 대장정을 알리는 성화가 막 점화되어 타오르고 있다.
- (숫자+회 말, 숫자+회 초) '3회 말', '9회 초' 등의 경우 '말, 초'를 제외하고 '3회, 9회'만 따

로 EV\_SPORTS로 태깅한다.

- 진짜 승부는 9회 말부터였다. 전반전, 후반전 풀타임을 소화한 손흥민 선수 → 진짜 승부는 <9회: EV\_SPORTS> 말부터였다. 전반전, 후반전, 풀타임을 소화한 손흥민 선수
- (중복 개체명) '유로'는 문맥을 준하여 중복 개체명 분류가 가능하다.
  - 유로 2008과 2012, 2010 남아공 월드컵을 제패해 축구 역사를 새로 쓴 스페인도 마찬가지. → <유로: EV\_SPORTS> 2008과 2012, 2010 남아공 월드컵을 제패해 축구 역사를 새로 쓴 스페인도 마찬가지.
  - 기존에 환전이 가능했던 미국 및 캐나다 달러와 유로, 엔화, 위안화에 홍콩 및 싱가포르 달러와 태국 바타를 추가해 환전 가능 통화를 늘렸다. → 기존에 환전이 가능했던 미국 및 캐나다 달러와 <유로: CV\_CURRENCY>, 엔화, 위안화에 홍콩 및 싱가포르 달러와 태국 바타를 추가해 환전 가능 통화를 늘렸다
- (개체명의 일부로 해당 개체를 가리키는 경우) 문맥을 고려하여 맥락상 EV\_SPORTS일 경우 태깅이 가능하다.
  - 평창 티켓 있으신가요? 20일 쇼트트랙 → <평창: EV\_SPORTS> 티켓 있으신가요? 20일 쇼트트랙
  - 88 서울 때 마스코트는 호돌이다. → <88 서울: EV\_SPORTS> 때 마스코트는 호돌이다.

## 11.4. EV\_FESTIVAL

### 11.4.1. 정의 및 설명

- 축제 명칭
  - 축제: '공연+퍼레이드'의 종합적 성격을 갖는 문화행사
- 영화제 명칭
- 콘서트 명칭
- 행사 명칭

### 11.4.2. 분석 대상

- 축제 명칭
  - 연어축제, 연등축제, 장미축제, 단풍축제, 갯방구대축제, 새미골막사발축제, 이천도자기축제, 수원화성문화제, 보령머드축제, 담양대나무축제, 부산바다축제, 부천국제만화축제, 울산서머페스티벌, 록 페스티벌
  - ~대(학교) 축제, ~대동제
- 영화제 명칭
  - 베니스 국제영화제, 칸 영화제, 부산국제영화제, 부천영화제
- 콘서트 명칭

- BOYSWISH 콘서트, 방탄 콘서트, 트와이스 콘서트, 나훈아 콘서트
- 행사 명칭
  - 책 읽는 춘천, 2019 서울모터쇼, 코리아 세일 페스타, 서울 패션 위크, 코베 베이비페어

#### 11.4.3. 예문

- 축제 명칭
  - 강원 양양군의 대표적 지역축제인 <연어축제: EV\_FESTIVAL>가 관광객 18만여 명이 다녀간 가운데 폐막했다.
  - 1973년 시작된 제등행진이 1996년 새롭게 단장되어 문화축제로 발전된 것이 <연등축제: EV\_FESTIVAL>이다.
  - 장미의 계절입니다. 곳곳에 <장미축제: EV\_FESTIVAL>입니다.
  - <이천도자기축제: EV\_FESTIVAL>는 우수한 이천도자기를 전세계에 널리 알리고, 우리 문화의 저변 확대와 지역 활성화를 위해 이천문화원 이천민속도자기사업협동조합 주관
  - <보령머드축제: EV\_FESTIVAL> 외국인들 참 많네요.
  - 20180510 <경희대학교 축제: EV\_FESTIVAL> 마마무 화사 직캠
  - '스트레스를 날려라' <성균관대 대동제: EV\_FESTIVAL> 개막
- 영화제 명칭
  - 바로 세계적인 영화제 중 하나, <베니스 국제 영화제: EV\_FESTIVAL>가 열렸었는데요!
  - 이번 주 <부산국제영화제: EV\_FESTIVAL> 개최한다고 한창 야외무대 준비중이네요.
- 콘서트 명칭
  - 세븐틴은 지난 13~14일 서울 올림픽공원 SK핸드볼경기장에서 데뷔 첫 앙코르 콘서트 <BOYSWISH 콘서트: EV\_FESTIVAL>를 개최했다.
  - 2019년 10월 27일, 29일 <방탄 콘서트: EV\_FESTIVAL>를 보고 왔다.
- 행사 명칭
  - 제67회 <책읽는 춘천: EV\_FESTIVAL> 일시: 2019년 05월 07일(화)
  - 다양한 신차와 볼거리로 가득한 2019 <서울모터쇼: EV\_FESTIVAL>로 당신을 초대합니다.
  - <서울패션위크: EV\_FESTIVAL>의 글로벌 위상을 강화하고 K뷰티·K팝·K스타일 등을 대표하는 행사로서 외연을 확장하겠습니다.
  - 일산 킨텍스에서 개최된 <코베 베이비페어: EV\_FESTIVAL>가 성황리에 종료됐다.

#### 11.4.4. 주의 사항

- ~축제, ~영화제, 행사 명칭을 대상으로 한다.
- EV\_FESTIVAL은 일반 유형도 태깅한다.
  - 연어축제, 장미축제, 연등축제, 단풍축제, 록 페스티벌 등 일반 유형의 경우도 포함한다.

## 11.5. EV\_OTHERS

### 11.5.1. 정의 및 설명

- 사건/사고 명칭
- ~사태
- 회의
- 학술대회(회의)
- 전시회, 영화 시사회, 팬사인회
- 경진대회
- 축제 내 체험행사

### 11.5.2. 분석 대상

- 사건/사고 명칭
  - 세월호 참사, 태안 기름유출사고, 헤이그특사 사건, 헤이그밀사 사건, 국정농단, 이메일 게이트, 플라자 합의, 포츠머스조약, 남북 평화협정
  - 일본 불매운동, 양담배 불매운동, 맥도날드 불매운동, 아이스버킷챌린지
- ~사태
  - 페르시아만사태, 한진중공업파업사태, 중동사태, 천안문사태, 동남아 쓰나미 사태
- 회의
  - G20 정상회의, 다보스포럼, 가이다르포럼, 퍼그워시 회의
- 학술대회(회의)
  - 국어학 국제학술대회, 국어 정책 학술 대회, 세계 여성학 대회, 국제 컴퓨터 통신 회의, 세계 정보 통신 국제 학술 회의, IEEE(국제전기전자기술자협회) 학술회의
- 전시회, 영화 시사회, 팬사인회
  - ART TOY CULTURE 2014 SEOUL, 샤갈 러브 앤 라이프전, 키스해링전
  - 체르노빌 상영회, 기생충 상영회, 어벤저스 상영회
  - 벡컴 팬사인회, 트와이스 팬사인회
- 경진대회
  - 국제뷰티기능대회, 청소년저작권글짓기대회, 해군호국문예제, 하나은행 자연사랑 어린이 미술대회, 미스유니버스, 월드바리스타챔피언십
- 축제 내 체험행사
  - 연어맨손잡기체험, 대나무체험행사, 카누체험, 북 쥐띠 문양 페이스페인팅

### 11.5.3. 예문

- 사건/사고 명칭
  - <세월호 참사: EV\_OTHERS>는 2014년 4월 16일 인천에서 제주로 향하던 중에 발생하였



- 다.
- 온 국민을 안타깝게 했던 <태안 기름유출사고: EV\_OTHERS>가 올해로 10년째가 되었다고 하네요.
  - 일본은 1907년, <헤이그 특사 사건: EV\_OTHERS>을 구실로 삼아 고종을 강제 퇴위시켰다.
  - 2019년 대한민국의 <일본 불매 운동: EV\_OTHERS>은 한일 무역 분쟁으로 인해서 본격적으로 촉발되었다.
  - 2014년 여름에 시작된 <아이스 버킷 챌린지: EV\_OTHERS>는 소셜 미디어를 통해 급격히 퍼져나가 하나의 유행이 되었다.
  - ~사태
    - 그런데 미국은 이번 <페르시아만 사태: EV\_OTHERS>를 자신들의 뜻대로 해결했을 경우, 소련을 국제무대에서 확실하게 따돌릴 수 있다고 확신하는 듯하다.
  - 회의
    - 文대통령의 2018 <G20 정상회의: EV\_OTHERS> 리트리트 세션 발언문 전문입니다.
    - 엘비나 나비올리나 러시아 중앙은행 총재가 1월 17일 모스크바에서 열린 <가이다르 포럼: EV\_OTHERS>에서 발언하고 있다.
    - 매년 <다보스포럼: EV\_OTHERS>에서는 '글로벌 지속가능경영 100대 기업'을 선정해 발표한다.
    - 50회를 맞은 올해 <다보스포럼: EV\_OTHERS>은 24일까지 나흘간 스위스 휴양지 다보스에서 진행됩니다.
    - 1957년 제1회 <퍼그워시 회의: EV\_OTHERS>가 열린 캐나다 퍼그워시 마을의 '생각하는 자의 오두막'.
  - 학술대회
    - 국립국어원은 10일 서울 중구 페럼타워에서 '2017년 <국어 정책 학술 대회: EV\_OTHERS>'를 연다고 밝혔다.
    - 24일까지 열리는 제9차 <세계여성학대회: EV\_OTHERS>에 참가한 세계적 여성학자들이 한국사회의 민감한 주제에 대해 입을 열었다.
    - 아파넷은 1972년 제1회 <국제 컴퓨터 통신 회의: EV\_OTHERS>에서 첫선을 보이며 세상에 모습을 드러냈는데요.
    - 디즈니는 7월 22일 하와이 호놀룰루에서 열린 <IEEE 학술회의: EV\_OTHERS>에서 관람객들의 표정을 실시간으로 분석해 영화 평점을 매기는 인공지능 '프베스(FVAEs)'를 개발했다고 밝혔다.
  - 전시회, 영화 시사회, 팬사인회
    - DDP에서 5월 1일~5일 전시되었던 <ART TOY CULTURE 2014 SEOUL: EV\_OTHERS>의 기록.
    - 예술은 삶, 삶은 곧 예술이라는 <키스해링겐: EV\_OTHERS> 11월 24일부터 3월 17일까

지 만나실수 있습니다.

- 전편 무료 <체르노빌 상영회: EV\_OTHERS>를 10월 20일 롯데시네마 월드타워 수퍼플렉스G에서 개최한다고 합니다.
- 29일 오후 <베컴 팬사인회: EV\_OTHERS>는 물려든 인파로 인한 안전상의 문제로 시작 15분 만에 취소됐다.
- 경진대회
  - 2019 <국제뷰티 기능대회: EV\_OTHERS> 출전 생생 후기 포스팅하려 합니다!
  - <월드바리스타챔피언십: EV\_OTHERS>이 오는 2017년 11월 9일부터 12일까지 개최된다.
- 축제 내 체험행사
  - 전국항에서 개최되는 2008 경기 보트쇼 및 세계요트대회 기간 중 체험행사의 일환으로 <바다낚시 체험: EV\_OTHERS>을 개최합니다.
  - 2008년 새해 국립민속박물관은 『무자년 설·대보름 맞이 대축제』에서 대형 쥐 문양판에 끼우는 <무자년 소원 담기 행사: EV\_OTHERS>를 비롯해 옷을 던져 새해 운수를 점치는 <옷점 보기: EV\_OTHERS>, <복 쥐띠 문양 페이스페인팅 프로그램: EV\_OTHERS> 등을 진행한다.
  - 2017 인사동 정월대보름행사는 희망찬 출발을 기원하는 내용의 <비나리 공연: EV\_OTHERS> 등 전통 국악공연을 선보입니다.

#### 11.5.4. 주의 사항

- EV\_OTHERS는 문맥상에서 일시, 장소가 명확한 특정 행사, 사건으로 볼 수 있는 항목인 경우에만 태깅한다.
- 날짜(DT)나 횟수(QT)가 함께 쓰일 때는 구분하여 태깅한다.
  - <1957년: DT\_YEAR> <제1회: QT\_ORDER> <퍼그워시 회의: EV\_OTHERS>가 열린 캐나다 퍼그워시 마을의 '생각하는 자의 오두막'.
- (선거, 주민투표) 문맥상에서 일시, 장소가 명확한 특정 행사, 사건으로 볼 수 있는 항목인 경우에만 태깅한다. '주민투표제'는 CV\_POLICY에 해당된다. 하지만 '주민 투표'는 '-시행하다, -진행하다' 등 적정 술어와 공기하며, 특정 행사, 사건에 해당하는 경우에는 EV\_OTHERS에 해당된다. '2011 서울특별시 무상급식 주민투표', '2005년 제주도 행정 구역 개편 주민투표', '대구 군 공항 이전 주민투표' 등 확실한 사건으로 볼 수 있는 경우만 태깅한다.
  - 지난 4월15일 수요일 대한민국의 제21대 총선이 있었습니다! → 지난 4월15일 수요일 대한민국의 제21대 <총선: EV\_OTHERS>이 있었습니다!
  - 2011년 8월에 오세훈 전 서울시장은 무상급식에 대한 주민투표를 통해 투표율 33.3%를 달성하지 못함에 따라 사퇴하였다. → 2011년 8월에 오세훈 전 서울시장은 무상급식에 대한 <주민투표: EV\_OTHERS>를 통해 투표율 33.3%를 달성하지 못함에 따라 사퇴하였다.

- (~정상회담, ~협상) EV\_OTHERS로 일시, 장소가 명확한 구체적인 사건으로 특정할 수 있는 경우에만 태깅한다.
  - '한일 서류없는 무역' 사업은 2000년 9월에 열린 한일 정상회담에서 채택된 '한일 IT협력 이니셔티브'의 후속조치로 시작됐다. → '한일 서류없는 무역' 사업은 2000년 9월에 열린 <한일 정상회담: EV\_OTHERS>에서 채택된 '한일 IT협력 이니셔티브'의 후속조치로 시작됐다.
  - 이스라엘과 팔레스타인 간의 중동평화협상이 1년 8개월만인 2일 워싱턴에서 공식 재개됐다. → 이스라엘과 팔레스타인 간의 <중동평화협상: EV\_OTHERS>이 1년 8개월만인 2일 워싱턴에서 공식 재개됐다.
- (수능, 임용고시 등 각종 선발 고사) 특정 시간에 특정 장소에서 이루어지는 사건일 경우에만 태깅한다. '-시행되다, -실시하다, -치르다' 등의 적정술어와 공기하는 경우 EV\_OTHERS로 태깅한다.
  - 3일 치러진 2021학년도 수능 3교시 영어 영역은 지난해 수능과 비슷한 수준으로 출제된 것으로 평가됐다. → 3일 치러진 2021학년도 <수능: EV\_OTHERS> 3교시 영어 영역은 지난해 수능과 비슷한 수준으로 출제된 것으로 평가됐다.
  - 올해 수능은 작년 수능에 비해 어렵다는 평이 있다. → 올해 <수능: EV\_OTHERS>은 작년 수능에 비해 어렵다는 평이 있다.
  - 지난해 수능 필적확인 문구는 박두진의 시 '별밭에 누워'에서 인용한 '너무 맑고 초롱한 그중 하나 별이여'였다. → 지난해 <수능: EV\_OTHERS> 필적확인 문구는 박두진의 시 '별밭에 누워'에서 인용한 '너무 맑고 초롱한 그중 하나 별이여'였다.
    - '올해 수능', '지난해 수능'에서 '수능'도 문맥상 특정 시간에서의 수능을 의미하므로 태깅한다.
  - 배우들도 멋있고 수능 공부로 쌓인 스트레스가 확 날아갈 만큼 역동적이고 재미있는 공연이었다.
    - 단지 입학 시험이나 선발 제도 등을 의미하면 태깅하지 않는다.
  - 2021 임용고시는 일정대로 11월 21일(토)에 시행했다. → 2021 <임용고시: EV\_OTHERS>는 일정대로 11월 21일(토)에 시행했다.
- (중복 개체명) '~조약', '~협정'은 문맥에 준하여 EV\_OTHERS와 CV\_LAW로 태깅한다. '-을 체결하다, -을 맺다' 등의 적정술어와 공기하는 경우이면 EV\_OTHERS로 태깅한다. '-에 쓰여져 있다, -에 서명하다' 등 조약서, 협정서를 의미하는 경우는 CV\_LAW로 태깅한다. 문맥에서 판단하기 어려울 경우에는 EV\_OTHERS를 우선순위로 정한다.
  - 1905년 9월 5일, 미국의 중재로 일본과 러시아가 미국에서 포츠머스 조약을 체결했다. → 1905년 9월 5일, 미국의 중재로 일본과 러시아가 미국에서 <포츠머스 조약: EV\_OTHERS>을 체결했다.
  - EU는 2011년 한국과 자유무역협정을 체결하면서 한국에 핵심협약 비준을 위해 노력해 달라고 요구했다. → EU는 2011년 한국과 <자유무역협정: EV\_OTHERS>을 체결하면서

한국에 핵심협약 비준을 위해 노력해 달라고 요구했다.

- 포츠머스 조약에는 러시아가 조선으로부터 철수하고 모든 영향력을 내려 놓겠다는 내용이 쓰여져 있다. → <포츠머스 조약: CV\_LAW>에는 러시아가 조선으로부터 철수하고 모든 영향력을 내려 놓겠다는 내용이 쓰여져 있다.

## 12. ANIMAL

---

- 생물은 2004년 스미스 분류체계에 따라 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, '사람'을 제외한 짐승을 대상으로 한다.
- 세분류에 포함되지 않은 동물명은 AM\_OTHERS로 태깅한다.
- 천연기념물로 지정된 동물명은 그에 해당하는 ANIMAL로 태깅한다.

### 12.1. AM\_INSECT

#### 12.1.1. 정의 및 설명

- 곤충의 이름
- 곤충은 절지동물에 속하나 따로 분류하며 머리-가슴-배의 구분이 있다.

#### 12.1.2. 분석 대상

- 곤충의 이름
  - 딱정벌레, 나비, 배추흰나비, 제왕나비, 매미, 참매미, 사마귀, 모기, 개미, 흰개미, 파리, 노린재, 벌, 말벌, 꿀벌, 호박벌, 무당벌레, 나방, 아틀라스나방, 풍뎅이, 잠자리, 소금쟁이 등

### 12.1.3. 예문

- 곤충의 이름
  - <딱정벌레: AM\_INSECT>는 어른 벌레 상태로 영하 60도까지 견딘다.
  - 특히 호주산 <나비: AM\_INSECT>들만 채집되어 있는 나비관(Butterfly House)은 멜번 동물원의 자랑거리이다.

## 12.2. AM\_BIRD

### 12.2.1. 정의 및 설명

- 새(조류)의 이름
- 새(조류)의 하위 유형

### 12.2.2. 분석 대상

- 새(조류)의 이름
  - 올빼미, 안경올빼미, 왜가리, 삿꾸기, 거위, 두루미, 황새, 참새, 딱따구리, 오리, 청둥오리, 타조, 기러기, 앵무새, 꾀꼬리, 까마귀, 까치, 펭귄, 페루펭귄, 닭, 벌새 등
- 새(조류)의 하위 유형
  - 철새, 겨울철새, 겨울새, 여름철새, 여름새, 텃새

### 12.2.3. 예문

- 새(조류)의 이름
  - <거위: AM\_BIRD>는 가금인 지 오래되었으므로 거친 먹이도 잘 먹고 추위에도 강하다.
  - 실제로 이 그림은 <두루미: AM\_BIRD>가 아니라 <황새: AM\_BIRD>가 잘못 그려진 것이라 한다.
- 새(조류)의 하위 유형 명칭
  - 세계적인 철새 도래지 천수만에 다양한 <겨울철새: AM\_BIRD>들이 몰려 장관을 이루고 있습니다.
  - 한강에서 가장 많은 종류의 <철새: AM\_BIRD>를 볼 수 있는 곳은 강서습지생태공원 일대.

### 12.2.4. 주의 사항

- (음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분한다. 조류가 동물의 의미일 경우는 '살다, 태어나다, 자라다, 날다, 낳다' 등의 적정술어와 같이 쓰인다.
  - 닭은 날개가 있으나 날지 못하는 조류이다. → <닭: AM\_BIRD>은 날개가 있으나 날지

- 못하는 조류이다.
- 오늘 잡은 닭으로 만든 백숙. → 오늘 잡은 <닭: CV\_FOOD>으로 만든 백숙.
- (새 이름+알) '새이름+알'의 경우 AM\_BIRD의 태깅 대상이 아니다.
  - 달걀, 계란, 메추리알 등은 AM\_BIRD로 태깅하지 않는다.

## 12.3. AM\_FISH

### 12.3.1. 정의 및 설명

- 어류의 이름
- 어류의 하위 유형 명칭

### 12.3.2. 분석 대상

- 어류의 이름
  - 참가자미, 장어, 갯장어, 붕장어, 전기뱀장어, 가물치, 고등어, 방어, 연어, 갈치, 멸치, 미꾸리, 송어, 날치, 붕어, 버들붕어, 꽂치, 돔, 돌돔, 참돔, 감성돔, 아귀, 초롱아귀 등
- 어류의 하위 유형 명칭
  - 민물고기, 담수어, 천어, 바닷고기, 바닷물고기, 해수어

### 12.3.3. 예문

- 어류의 이름
  - 《난호어목지》에서는 <갈치: AM\_FISH>라 하였다.
  - 옛부터 식용과 약용으로 널리 쓰인 <미꾸리: AM\_FISH>는 지금도 맑은 물에 담귀 놓아 오물을 모두 토하게 한 후, 꼭 고아서 국을 끓인 추어탕의 원재료이다.
  - <송어: AM\_FISH>는 예로부터 음식으로서만 아니라 약재로도 귀하게 여겼다.
  - <날치: AM\_FISH> 무리 중에는 열대성 <날치: AM\_FISH>와 아열대성 <날치: AM\_FISH>, 온대성 <날치: AM\_FISH> 등이 있다.
- 어류의 하위 유형 명칭
  - 18일 정취안르바오(證券日報)가 중국 수산양식망을 인용해 보도한 바에 따르면 잉어와 연어 붕어 등 <민물 고기: AM\_FISH> 가격이 각각 7.45%, 5.36%, 4.47% 씩 상승했다.

### 12.3.4. 주의 사항

- (음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분한다. 어류가 동물의 의미일 경우는 '살다, 태어나다, 자라다, 헤엄치다, 낚다' 등의 적정술어와 같이 쓰인다.
  - 고등어는 따뜻한 바다에서 서식한다. → <고등어: AM\_FISH>는 따뜻한 바다에서 서식한다.

- 오늘 저녁에는 고등어로 요리를 해야겠다. → 오늘 저녁에는 <고등어: CV\_FOOD>로 요리를 해야겠다.
- (어류 이름+알) '어류 이름+알'의 경우 AM\_FISH의 태깅 대상이 아니다.
  - 청어알, 상어알, 날치알 등은 AM\_FISH로 태깅하지 않는다.

## 12.4. AM\_MAMMALIA

### 12.4.1. 정의 및 설명

- 포유류의 이름

### 12.4.2. 분석 대상

- 포유류의 이름
  - 개, 고양이, 토끼, 고래, 호랑이, 코카스파니엘, 벵갈고양이, 스코티쉬폴드, 은여우토끼, 범고래, 너구리, 오리너구리, 박쥐, 사슴, 붉은사슴, 캥거루, 고릴라, 기린, 다람쥐, 코알라 등

### 12.4.3. 예문

- 포유류의 이름
  - 일반적으로 몸길이 4~5m 이상인 것을 <고래: AM\_MAMMALIA>, 소형인 것을 <돌고래: AM\_MAMMALIA>라고 하는데, <망치고래: AM\_MAMMALIA>와 같이 소형인 것도 <고래: AM\_MAMMALIA>라는 이름이 붙은 종류가 있다.
  - 겨울이 되면 일부 이주하는 <박쥐: AM\_MAMMALIA>를 제외하고 대부분이 겨울잠을 잔다.
  - <코카스파니엘: AM\_MAMMALIA>은 사람들과 장난치기를 좋아하며, 훈련 적응력도 높다.
  - 호랑이 중 가장 북방 지역에 서식한다는 특성 상 <시베리아호랑이: AM\_MAMMALIA>는 겨울에는 겨울털이 길게 자라고 배와 옆구리에 두터운 지방층이 있어 배 부분이 늘어져 보인다.

### 12.4.4. 주의 사항

- (동물의 이름) 사람이 호명하기 위해 붙인 동물의 이름은 PS\_PET으로 태깅한다.
  - 우리집 개 누렁이는 용맹해서 집을 잘 지킨다. → 우리집 <개: AM\_MAMMALIA> <누렁이: PS\_PET>는 용맹해서 집을 잘 지킨다.
- (인간) 사람, 인간 및 인류 진화의 개체는 태깅하지 않는다.
  - '인간, 사람, 호모사피엔스, 오스트랄로피테쿠스' 등은 태깅하지 않는다.
- (신조어) 해당 포유류의 특성을 구체화하는 '개냥이, 똥냥이' 등과 귀엽게 부르는 애칭인 '

멍멍이, 땡땡이' 등도 AM\_MAMMALIA로 태깅한다.

- 스코티쉬폴드는 개냥이로 유명하다. → <스코티쉬폴드: AM\_MAMMALIA>는 <개냥이: AM\_MAMMALIA>로 유명하다.

## 12.5. AM\_AMPHIBIA

### 12.5.1. 정의 및 설명

- 양서류의 이름

### 12.5.2. 분석 대상

- 양서류의 이름
  - 개구리, 두꺼비, 황소개구리, 산청개구리, 도롱뇽, 꼬마도롱뇽, 꼬리치레도롱뇽, 맹꽁이, 올챙이 등

### 12.5.3. 예문

- 양서류의 이름
  - 닭머리만 꼬치에 꿰인 것도 있었고 <황소개구리: AM\_AMPHIBIA>도 즐겨 먹는다고 합니다.
  - 맑고 깨끗한 물에서 물고기나 <개구리: AM\_AMPHIBIA>를 먹고 사는 천연기념물.
  - 한국에는 <도롱뇽: AM\_AMPHIBIA>(<Hynobius leechi: AM\_AMPHIBIA>)-<꼬리치레도롱뇽: AM\_AMPHIBIA>(<Onychodactylusfischeri: AM\_AMPHIBIA>)-<네발가락도롱뇽: AM\_AMPHIBIA>(<Salamandrella keyserlingi: AM\_AMPHIBIA>)의 3 속 3 종이 있으며,

## 12.6. AM\_REPTILIA

### 12.6.1. 정의 및 설명

- 파충류의 이름

### 12.6.2. 분석 대상

- 파충류의 이름
  - 도마뱀, 왕도마뱀, 거북, 자라, 뱀, 방울뱀, 살모사, 구렁이, 백사, 공룡, 익룡, 디노사우르스, 스피노 사우르스, 악어, 미국악어, 인도악어, 이구아나, 초록이구아나 등

### 12.6.3. 예문



- 파충류의 이름
  - 우리 나라에는 <구렁이: AM\_REPTILIA>에 관한 전설과 전승 속신이 많이 전한다.
  - 흔히 민간이나 한방에서는 <백사: AM\_REPTILIA>를 희귀한 보약으로 여기고 있으나, 과학적으로는 근거가 없다.
  - <악어: AM\_REPTILIA>는 담수성이나, <인도악어: AM\_REPTILIA>의 경우에는 바다에도 나온다.
  - 한데 묶어 <디노사우르: AM\_REPTILIA>라고 한 것을 동양에서 '<공룡: AM\_REPTILIA>'이라고 번역하면서 쓰이기 시작했다.

## 12.7. AM\_TYPE

### 12.7.1. 정의 및 설명

- 동물 분류의 명칭

### 12.7.2. 분석 대상

- 동물 분류의 명칭
  - 강장동물, 구두동물, 극피동물, 내항동물, 모악동물, 선형동물, 성구동물, 연체동물, 완보동물, 완족동물, 원생동물, 유선형동물, 유형동물, 윤형동물, 의충동물, 절지동물, 척색동물, 태형동물, 편형동물, 해면동물, 환형동물, 포유류, 메뚜기과, 어류, 조류, 양서류, 파충류

### 12.7.3. 예문

- 동물 분류의 명칭
  - 200 여 종의 <조류: AM\_TYPE>와 뱀, 개구리, 도마뱀 등 <파충류: AM\_TYPE>가 서식하고 있다.
  - 실제 크기의 바다 <양서류: AM\_TYPE>와 공룡 화석, 호주에서 가장 큰 나방이 볼 만하다.

### 12.7.4. 주의 사항

- (새/물고기) 조류와 어류를 통칭할 수 있는 '새', '물고기'는 AM\_TYPE으로 태깅한다.
  - 제작진은 불과 400년 전까지 지구상에 존재했던 가장 큰 새의 흔적을 찾기로 했다. → 제작진은 불과 400년 전까지 지구상에 존재했던 가장 큰 <새: AM\_TYPE>의 흔적을 찾기로 했다.

## 12.8. AM\_PART

### 12.8.1. 정의 및 설명

- 동물이나 사람 몸의 한 부분(신체 부위) 명칭
- 일반적으로 눈으로 확인할 수 있는 몸의명칭
  - 사람 신체
  - 동물 몸

### 12.8.2. 분석 대상

- 일반적으로 눈으로 확인할 수 있는 몸의 명칭
  - 사람 신체: 팔, 다리, 머리, 장단지, 허벅지, 팔꿈치, 눈, 코, 입, 이, 혀, 머리카락, 눈썹, 피부
  - 동물 몸: 팔, 다리, 머리, 대가리, 지느러미, 부리, 날개, 꼬리, 이빨, 혀, 털, 피부

### 12.8.3. 예문

- 사람 신체
  - 경기도 고양경찰서는 지난 11일 한 학원 건물 입구에서 14살 김 모양이 남자아이 2명의 <다리: AM\_PART>를 걸어 넘어뜨리는 영상이 공개되어 오늘 오전 김 양을 불러 경위를 파악할 계획이라고 밝혔습니다 .
  - 그래서 아마도 우리의 컴퓨터가 사람의 <머리: AM\_PART>하고 비슷한 일을 머지 않아서 할 것이다, 그런 생각을 과학자들은 하고 있죠.
  - 우리 <몸: AM\_PART> 깊은 속에서부터 작용을 하게 합니다.
- 동물 몸
  - 갯가지 화려한 색상을 뽐어내는 산호초와, <지느러미: AM\_PART>와 <부리: AM\_PART>가 용의 모습을 닮은 해룡 (Sea-Dragon), 표범무늬가 독특한 바다뱀 (Sea Snake)은 반드시 보도록 한다.

### 12.8.4. 주의 사항

- (뼈) 눈으로 확인할 수 있는 명칭만 해당된다.
  - 골반이 틀어졌다. → <골반: TM\_CELL\_TISSUE\_ORGAN>이 틀어졌다.
    - '골반'은 뼈를 지칭하는 것이기 때문에 AM\_PART로 태깅하지 않는다.
- (껍질) '껍질'은 물질을 나타내므로, AM\_PART에 해당하지 않는다.
- (살) '희다, 굵다, 타다' 등의 술어와 쓰이는 '피부'의 의미를 지닌 '살'만 태깅한다.
  - 햇볕에 살이 많이 탔다. → 햇볕에 <살: AM\_PART>이 많이 탔다.
  - 배에 살이 많이 찼다. → <배: AM\_PART>에 살이 많이 찼다.
    - 이때 '살'은 '피부'를 의미하는 것이 아니기 때문에 태깅하지 않는다. 이와 마찬가지로 '뱃살, 팔뚝살' 등은 태깅하지 않는다.
- (조직, 기관) 신체 조직이나 기관 중에서 눈으로 보이는 대상은 TM\_CELL\_TISSUE\_ORGAN이

아닌 AM\_PART로 태깅한다.

- 올바른 치솔질로 이를 닦아야 합니다. → 올바른 치솔질로 <이: AM\_PART>를 닦아야 합니다.
- 강아지의 부드러운 털을 쓰다듬으며 → 강아지의 부드러운 <털: AM\_PART>을 쓰다듬으며
- (방향+신체부위) '방향+신체부위'는 따로 태깅하지 않고 통합하여 태깅한다.
  - 왼발을 오른 다리 위에 놓고 → <왼발: AM\_PART>을 <오른 다리: AM\_PART> 위에 놓고
- (접두사+신체 부위) 신체 부위에 접두사가 함께 쓰인 '맨다리, 맨몸, 맨눈, 맨발' 등의 경우도 함께 태깅한다.
  - 개기일식은 맨눈으로 보면 위험하다. → 개기일식은 <맨눈: AM\_PART>으로 보면 위험하다.
- (신체 부위+국소지칭) 특정 신체부위의 일부분을 지칭하는 '입꼬리, 손끝, 발끝' 등과 같은 경우는 태깅하지 않는다.
  - 쇼파 밑에 있는 돈을 손끝으로 겨우 꺼냈다.
    - '손끝'은 태깅하지 않고 '손'도 따로 태깅하지 않는다.

## 12.9. AM\_OTHERS

### 12.9.1. 정의 및 설명

- 기타 동물 명칭
- 기타 동물 명칭의 하위 부류
  - 절지동물류
  - 연체동물류
  - 기타 생물류
  - 미생물류
  - 상상 속 동물류

### 12.9.2. 분석 대상

- 절지동물류
  - 가재, 새우, 대하, 토하, 게, 대게, 꽃게, 지네
- 연체동물류
  - 오징어, 문어, 낙지, 달팽이, 민달팽이, 조개, 굴, 전복
- 기타 생물류
  - 해파리, 산호, 멧게, 해삼, 불가사리, 말미잘, 지렁이

- 미생물류
  - 세균, 바이러스, 박테리아, 아메바
- 상상 속 동물류
  - 유니콘, 용, 해태, 봉황, 삼죽오

### 12.9.3. 예문

- 절지동물류
  - <가재: AM\_OTHERS>는 <게: AM\_OTHERS>편이라고, 감사합니다.
  - <붉은대게: AM\_OTHERS>는 1,000m 바다 밑 청정지역에 삽니다.
- 연체동물류
  - <문어: AM\_OTHERS>가 틀렸다고 해서 <오징어: AM\_OTHERS>로 해서
  - 유통업계에 따르면 24일부터 북한산 <조개: AM\_OTHERS> 반입이 중단된 이후 이마트에서 국산 조개류 가격이 10% 가량 상승한 것으로 나타났습니다.
- 기타 생물류
  - 슬리피 라군이라 불리는 <산호초: AM\_OTHERS>들이 파도를 막아 주면서 자연적으로 형성된 이스트아가나 비치에 있는 호텔이다.
- 미생물류
  - 뛰어난 항균력으로 <곰팡이: AM\_OTHERS>나 <박테리아: AM\_OTHERS>는 물론 오염에도 강한 내구성이 좋은 소재예요!
- 상상 속 동물류
  - 드래곤보트 위에서는 그들 모두가 다 광저우의 <용: AM\_OTHERS> 아닐까요?

### 12.9.4. 주의 사항

- (음식) 개체가 음식으로 쓰였을 때는 CV\_FOOD로 태깅한다. 이때는 '-가 맛있다' '-를 먹다', '-로 -를 만들다', '-를 요리하다', '-로 조리하다' 등의 적정 술어로 판단할 수 있다.
  - 민물 새우를 토하라고 합니다. → <민물 새우: AM\_OTHERS>를 <토하: AM\_OTHERS>라고 합니다.
  - 대하, 맛있죠. → <대하: CV\_FOOD>, 맛있죠.
- (미생물류) 미생물류는 문맥에 따라 태깅한다.
  - 김치 유산균을 배양해 만든 소금으로 → 김치 <유산균: AM\_OTHERS>을 배양해 만든 소금으로
  - 유산균이 장 건강에 좋다. → <유산균: TMM\_DRUG>이 장 건강에 좋다.
    - 영양보조제의 의미일 경우 TMM\_DRUG으로 태깅한다.
  - 노로바이러스를 예방하기 위해서 → <노로바이러스: TMM\_DISEASE>를 예방하기 위해서
    - 질병의 원인의 의미일 경우 TMM\_DISEASE로 태깅한다.

## 13. PLANT

---

PLANT는 생물계에서 동물과 세균을 제외한 꽃과 나무, 풀과 같은 육지식물, 해조류, 버섯류, 이끼류 등을 대상으로 한다.

### 13.1. PT\_FRUIT

#### 13.1.1. 정의 및 설명

- 과일을 포함한 식물의 열매나 씨앗

#### 13.1.2. 분석 대상

- 토종과일
  - 감, 구기자, 머루, 모과, 무화과, 산사자, 산수유, 오미자 등
- 열대과일
  - 바나나, 파인애플, 아보카도, 키위, 망고, 망고스틴, 코코넛, 두리안, 롱안, 리이치, 패션후르츠 등
- 굴류 과일
  - 귤, 감귤, 레몬, 라임, 오렌지, 자몽, 유자 등
- 딸기류 과일
  - 블루베리, 체리, 크랜베리, 라즈베리, 복분자딸기, 서양복분자, 딸기, 곰딸기, 멧딸기 등
- 핵과류 과일
  - 대추, 매실, 복숭아, 살구, 앵두, 자두 등
- 견과류 열매
  - 땅콩, 밤, 아몬드, 잣, 피스타치오, 헤즐넛, 호두 등
- 곡류의 열매
  - 옥수수, 쌀, 조, 수수, 벼, 보리, 밀, 기장, 귀리, 메밀, 퀴노아 등
- 콩류 열매
  - 완두콩, 팥, 녹두, 강낭콩, 동부, 서리태, 검은콩, 약콩, 쥐눈이콩 등
- 기타 과일 및 식물의 열매나 씨앗
  - 사과, 포도, 배, 수박, 참외, 후추, 멜론 등

#### 13.1.3. 예문

- 진짜 <감: PT\_FRUIT> 색깔이 참 예쁩니다.
- 특히 자연지리적으로 과수 재배에 적합하여 <복숭아: PT\_FRUIT>가 많이 생산되며
- <오렌지: PT\_FRUIT>의 윗면을 잘라 속을 파낸다.
- 한국에는 고려시대에 중국에서 들여온 것으로 추측되며 《조선왕조실록》 등에도 <포도: PT\_FRUIT>에 관한 기록이 실려 있다.
- <호두: PT\_FRUIT>는 열매가 성숙된 가을에 따서 물에 오랫동안 담가 두거나 한 자리에 쌓아 두어 썩힌 육질의 외과피를 제거하고 햇볕에 말린 뒤 딱딱한 내과피를 깨서 종자를 취한다.
- <메밀: PT\_FRUIT>은 서늘하고 습한 기후에서 잘 자란다.
- <후추: PT\_FRUIT>는 10~12월에 구형의 붉은 열매가 열린다.
- 논에 <벼: PT\_FRUIT>의 열매가 누렇게 익어간다.
- 고온다습한 기후에서 잘 자라는 <쌀: PT\_FRUIT>은 온대나 열대 기후에서만 재배 가능하다.

#### 13.1.4. 주의 사항

- (곡류의 열매) 옥수수와 같은 곡류를 PT\_FRUIT로 태깅한다.
  - 밭에 심은 옥수수가 익어가고 있다. → 밭에 심은 <옥수수: PT\_FRUIT>가 익어가고 있다.
- (음식VS식물의 열매) 쌀과 귀리와 같은 곡류의 열매를 가공하거나 요리한 밥, 죽, 귀리가루, 볶은 귀리, 현미 등은 PT\_FRUIT가 아니라 CV\_FOOD로 태깅한다.
  - 귀리는 벼과 식물이며 식용이나 사료용으로 사용된다. → <귀리: PT\_FRUIT>는 벼과 식물이며 식용이나 사료용으로 사용된다.
- 산세가 험악한 산지에 귀리를 심었다. → 산세가 험악한 산지에 <귀리: PT\_FRUIT>를 심었다.
  - 매일 아침 볶은 귀리를 우유에 타 먹었다. → 매일 아침 볶은 <귀리: CV\_FOOD>를 우유에 타 먹었다.
- (식물의 열매) '익다, 여물다, 맺다'와 같은 술어와 함께 쓰이면 완두콩과 같은 콩의 부류를 PT\_FRUIT로 태깅한다.
  - 꼬투리마다 완두콩이 풍성하게 여물었다. → 꼬투리마다 <완두콩: PT\_FRUIT>이 풍성하게 여물었다.
- (복합어 내부 태깅) 다음과 같은 복합어 내부의 구성요소는 PT\_FRUIT로 태깅하지 않는다. <우리말샘>에 '-' 또는 '^' 기호와 함께 등재되어 있는 경우, 복합어에 준하는 것으로 보아 내부 요소인 과일명을 태깅하지 않는다.
  - 자몽내, 사과내, 사과취, 굴껍질, 사과껍질 등
- (과일명+향) <우리말샘>에 구 단위로 등재되어 있는 경우, 즉 띄어쓰기가 되어 있는 경우는 한 단어(복합어)가 아니므로, 명사구 전체는 태깅하지 않지만, 내부 요소인 과일명은 PT\_FRUIT로 태깅한다.

- 사과 향 → <사과: PT\_FRUIT> 향
- 오렌지 향 → <오렌지: PT\_FRUIT> 향
- (과일명+추출물) 망고 추출물과 석류 추출물의 경우에는 망고, 석류와 같은 계열의 단어인 치자가 '치자 추출물'로서 <우리말샘>에 구 단위로 등재되어 있으므로, 이에 준하여 과일 명만 PT\_FRUIT로 태깅한다.
  - 치자 추출물 → <치자: PT\_FRUIT> 추출물
  - 망고 추출물 → <망고: PT\_FRUIT> 추출물
  - 이 음료에는 석류 추출물이 들어 있다. → 이 음료에는 <석류: PT\_FRUIT> 추출물이 들어 있다.
- (견과류 열매) 호두와 같은 견과류 열매는 PT\_FRUIT로 태깅한다.
  - <호두: PT\_FRUIT>는 호두나무의 열매로서 속살은 지방이 많고 맛이 고소하여 식용하며, 한방에서 변비나 기침의 치료, 동독(銅毒)의 해독 따위의 약재로 쓴다.
  - 호두가 맛있게 익었다. → <호두: PT\_FRUIT>가 맛있게 익었다.

## 13.2. PT\_FLOWER

### 13.2.1. 정의 및 설명

- 꽃 이름

### 13.2.2. 분석 대상

- 봄꽃
  - 개나리, 진달래, 제비꽃, 짙레꽃, 벚꽃, 튜립, 철쭉, 목련, 유채꽃, 민들레, 산수유꽃, 할미꽃 등
- 여름꽃
  - 장미, 목향장미, 금계국, 나팔꽃, 해바라기, 무궁화, 연꽃, 봉숭아, 접시꽃, 수국, 초롱꽃, 바람꽃, 금잔화, 금불초, 들국화, 수레국화, 가는동자꽃, 가래바람꽃 등
- 가을꽃
  - 갯쑥부쟁이, 국화, 겹삼입국화, 감국, 산국, 코스모스, 억새꽃, 메밀꽃, 백일홍, 채송화 등
- 겨울꽃
  - 동백, 수선화, 군자란, 프리지아, 칼랑코에, 복수초 등
- 그 외 꽃 이름
  - 데이지, 안개꽃, 백합, 히아신스, 개꽃 등

### 13.2.3. 예문

- <제비꽃: PT\_FLOWER>은 성실과 겸손을 나타낸다고 하였다.

- <히아신스: PT\_FLOWER>는 화단뿐만 아니라 분재나 물재배로도 많이 이용된다.
- 화관은 길이 4~5cm 이고 초롱(호롱)같이 생겨 <초롱꽃: PT\_FLOWER>이라고 한다.
- 흰 털로 덮인 열매의 덩어리가 할머니의 하얀 머리카락같이 보이기 때문에 <할미꽃: PT\_FLOWER>이라는 이름이 붙었다.
- 마당에 만개한 <백일홍: PT\_FLOWER>을 보며 친구와 커피를 마셨다.

#### 13.2.4. 주의 사항

- (꽃 이름) '구체적인 꽃의 명칭'이나 '구체적인 꽃의 명칭 + 꽃'일 경우에 PT\_FLOWER로 태깅한다. 구체적인 꽃의 명칭이 꼭 있어야 한다.
  - 길에 핀 민들레가 밟혀도 다시 일어났다. → 길에 핀 <민들레: PT\_FLOWER>가 밟혀도 다시 일어났다.
- (꽃 이름+꽃) 구체적인 꽃 이름 명칭 뒤에 '~꽃(예: 백합꽃)'이 붙어서 나온 경우 통합하여 태깅한다.
  - 제라늄 → <제라늄: PT\_FLOWER>
  - 제라늄꽃 → <제라늄꽃: PT\_FLOWER>
- (꽃 이름) '피다, 지다, 만발하다, 개화하다, 시들다, 피우다'와 같은 적정술어와 같이 쓰이면 '해바라기'와 같은 꽃의 부류를 PT\_FLOWER로 태깅한다.
  - 큰 키의 해바라기들이 밭에 만발하게 피어있다. → 큰 키의 <해바라기: PT\_FLOWER>들이 밭에 만발하게 피어있다.
- (~꽃) '나팔꽃'과 같이 음식물이나 다른 태깅 범주와 혼동될 여지가 없이 '꽃 이름'으로서 명확한 경우에는 적정술어가 없어도 PT\_FLOWER로 태깅한다.
- (다의어) 식물명은 식물의 일부분을 가리키기도 하고 식물 전체를 가리키기도 하는 등 다의성을 가지는데, 적정술어를 이용해서 하위 의미를 구분하여 태깅한다.
  - 산수유가 노랗게 피었다. → <산수유: PT\_FLOWER>가 노랗게 피었다.
  - 빨강게 익은 산수유를 따다. → 빨강게 익은 <산수유: PT\_FRUIT>를 따다.
  - 마당에 산수유를 심었다. → 마당에 <산수유: PT\_TREE>를 심었다.

### 13.3. PT\_TREE

#### 13.3.1. 정의 및 설명

- 나무 이름

#### 13.3.2. 분석 대상

- 침엽수
  - 삼나무, 소나무, 잣나무, 전나무, 분비나무, 가문비나무, 종비나무, 잎갈나무, 주목, 편백나



무, 낙엽송, 낙우송, 눈잣나무, 눈측백나무, 솔송나무, 향나무, 구상나무, 풍산가문비나무 등

- 활엽수

- 참나무, 벃나무, 호도나무, 물푸레나무, 피칸, 오리나무, 느티나무, 단풍나무, 목련나무, 피나무, 오동나무, 떡갈나무, 뽕나무, 상수리나무, 동백나무, 너도밤나무 등

### 13.3.3. 예문

- 한국에서는 <후박나무: PT\_TREE> 숲을 생활환경으로 하고 있어서 <후박나무: PT\_TREE>와 흑비둘기의 분포가 밀접한 관계를 가지고 있으며
- 부인 일가는 목은 <벃나무: PT\_TREE>들이 짝혀 넘어가는 소리를 들으면서
- 해발고도 1,000~1,200m 위에는 <분비나무: PT\_TREE> · <가문비나무: PT\_TREE> 등의 아한대성 침엽수림과 활엽수림의 혼합림이 분포한다.
- 일제침탈과 한국전쟁으로 황폐화된 산림을 복구하기 위해 산림청에서는 <백합나무: PT\_TREE>, <편백나무: PT\_TREE>, <소나무: PT\_TREE> 등 여러 종의 나무를 대규모로 심는 산림녹화사업에 힘써 왔다.

### 13.3.4. 주의 사항

- (나무류) 가지치기하다, 남벌하다, 베다, 자르다, 심다' 등의 적정술어와 같이 쓰일 경우에 '계피나무'와 같은 나무류를 PT\_TREE로 태깅한다.
  - 계피나무를 가지치기하다. → <계피나무: PT\_TREE>를 가지치기하다.
  - 은행나무는 우리나라의 기후와 토질에서 잘 성장하며 병충해 예방에도 도움이 되기 때문에 전국 지자체에서 가로수로 많이 심고 있다. → <은행나무: PT\_TREE>는 우리나라의 기후와 토질에서 잘 성장하며 병충해 예방에도 도움이 되기 때문에 전국 지자체에서 가로수로 많이 심고 있다.
- (다의성) 식물명이 나무를 가리키기도 하고 꽃을 가리키기도 하는 경우, 적정술어에 따라 구분하여 태깅한다.
  - 무궁화를 심었다. → <무궁화: PT\_TREE>를 심었다.
  - 무궁화가 피었다. → <무궁화: PT\_FLOWER>가 피었다.
- (나무 이름) '소나무'와 같이 음식물이나 다른 태깅 범주와 혼동될 여지가 거의 없고 '나무 이름'이 명확한 경우에는 적정술어가 없어도 PT\_TREE로 태깅한다.
- (천연기념물) '양평 용문사 은행나무'와 같은 천연기념물도 PT\_TREE로 태깅한다.
  - 양평 용문사 은행나무 → 양평 용문사 <은행나무: PT\_TREE>
- (현상) '단풍'이 나무를 가리킬 때에는 PT\_TREE로 태깅하지만, '단풍'이 '늦가을에 식물의 잎이 붉은빛이나 누런빛으로 변하는 현상'의 의미를 나타낼 때에는 태깅하지 않는다.
  - 요즘은 전원주택에 단풍을 많이 심는다. → 요즘은 전원주택에 <단풍: PT\_TREE>을 많이 심는다.

- 설악산에 단풍을 보러 갔다.
- 온 산에 빨갛고 노랗게 단풍이 들었다.

## 13.4. PT\_GRASS

### 13.4.1. 정의 및 설명

- 채소를 포함한 풀 이름

### 13.4.2. 분석 대상

- 엽채류
  - 갓, 근대, 머위, 무청, 미나리, 배추, 상추, 샐러리, 시금치, 쑥갓, 양배추 등
- 경채류
  - 두릅, 마늘쭉, 부추, 샐러리, 아스파라거스, 연교, 죽순, 파 등
- 근채류
  - 당근, 래디쉬, 무, 비트, 생강, 순무, 연근, 우엉, 토란 등
- 구근류
  - 마늘, 샬롯, 양파, 감자, 고구마 등
- 과채류
  - 가지, 고추, 동아, 박, 오이, 월과, 피망, 토마토, 호박 등
- 화채류
  - 브로콜리, 아티초크, 양하, 컬리플라워 등
- 기타 풀 이름
  - 토끼풀, 향부자, 하늘타리, 갈대, 질경이, 강아지풀, 창포, 억새, 잔디, 담배, 파슬리, 한란 등

### 13.4.3. 예문

- 판이 이 <갈대: PT\_GRASS>를 꺾어 피리를 만들어 그녀를 그리워하며 불었던 데서 <갈대: PT\_GRASS>를 음악의 상징으로 여기게 되었다.
- <수강아지풀: PT\_GRASS>(var. gigantea)은 조와 <강아지풀: PT\_GRASS>의 잡종이다.
- 뿌리줄기를 <창포: PT\_GRASS>라 한다.
- 이거 완전 <인삼: PT\_GRASS> 뿌리같이 생겼는데요.
- <고구마: PT\_GRASS> 수확이 오늘 미션입니다.
- 화분에 심은 <파슬리: PT\_GRASS>는 매일 햇빛을 쬌도록 해 줘야 잘 자란다.
- <한란: PT\_GRASS>은 한국에서는 제주도의 한라산에서 자라며, 천연기념물로서 보호받고 있다.

#### 13.4.4. 주의 사항

- (채소류 풀) '뽕다, 베다, 캐다'와 같은 술어와 함께 쓰이면 양파, 토마토, 감자, 시금치, 생강과 같은 채소류를 PT\_GRASS로 태깅한다.
  - 감자를 밭에서 캐서 바구니에 수북히 담았다. → <감자: PT\_GRASS>를 밭에서 캐서 바구니에 수북히 담았다.
- (음식VS채소류 풀) 문맥상 음식 재료명으로 쓰였다고 판단되는 경우, PT에 속하는 개체라 할지라도 CV로 태깅한다. 무조건적으로 PT를 우선시하여 태깅하지 않는다. 문맥에 따라 다르게 태깅한다.
  - 그런데 <양배추: CV\_FOOD>를 조금 먹어선 크게 효과를 볼 수 없어요. 그래도 <양배추: PT\_GRASS>는 좋으니 일상생활에서 자주 먹으려고 합니다. '그래도 양배추는 좋으니'에서 양배추는 PT\_GRASS로 태깅하나, CV\_FOOD로도 태깅 가능하다.
  - 채소의 경우, CV\_FOOD인지, PT\_GRASS인지 판단이 어려울 경우가 있는데 문맥을 최우선적으로 고려하며, 식물을 요리하거나 가공한 경우, 음식으로 먹는 경우에는 CV\_FOOD로 태깅한다.
- (복합어 내부 태깅) 감자밭은 <우리말샘>에 등재되어 있어 내부 태깅하지 않는다.
  - 딸기밭은 <우리말샘>에 등재되어 있지 않으나, 감자밭과 같이 처리하여 내부 태깅하지 않는다.
- (담배) 담배 (비추천명: 연초)는 문맥에서 명확하게 식물, 담뱃잎을 의미하는 경우에 해당되었을 때만 PT\_GRASS로 태깅한다. 그 외의 경우에는 태깅하지 않는다. 담배 상품명이나 경우에는 AFW\_OTHER\_PRODUCTS로 태깅한다.
  - 올해는 장마로 담배농사를 망쳤다. → 올해는 장마로 <담배: PT\_GRASS> 농사를 망쳤다.
  - 벌레가 담배의 잎을 갉아먹었다. → 벌레가 <담배: PT\_GRASS>의 잎을 갉아먹었다.

### 13.5. PT\_TYPE

#### 13.5.1. 정의 및 설명

- 식물 유형 명칭, 식물의 세부 종을 구분하는 유형

#### 13.5.2. 분석 대상

- 식물 유형, 세부 종 명칭
  - 녹조류, 선태식물, 양치식물, 양치류, 종자식물
  - 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
  - 콩과, 참나무과, 생강과, 박과

#### 13.5.3. 예문

- <박과: PT\_TYPE>의 덩굴성 식물인 동아는 손부가 직접 기르고 수확한 것입니다.

#### 13.5.4. 주의 사항

- (식물유형 및 세부 종) 다음은 PT\_TYPE으로 태깅하지 않는다.
  - 저지대식물, 고산식물, 꽃, 나무, 열대과일, 야생화 등
  - 과실류, 견과류, 곡류, 버섯류 등
- (식물유형 및 세부 종) 다음과 같은 경우 전체를 하나로 태깅하지 않고 각각 따로 태깅한다.
  - 담자균문 구멍장이버섯목 구멍장이버섯과 조개껍질버섯속 → <담자균문: PT\_TYPE> <구멍장이버섯목: PT\_TYPE> <구멍장이버섯과 : PT\_TYPE> <조개껍질버섯속: PT\_TYPE>

### 13.6. PT\_PART

#### 13.6.1. 정의 및 설명

- 눈으로 확인 가능한 식물의 한 조직, 부분에 대한 명칭

#### 13.6.2. 분석 대상

- 식물의 한 조직이나 부분 명칭
  - 싹, 나무갓수관, 뿌리, 뿌리목, 얇은 뿌리, 겉뿌리, 뿌리털
  - 가지, 큰가지, 곁가지, 잔가지, 가시, 열매, 잎, 줄기, 밑줄기
  - 꽃, 꽃잎, 꽃봉오리, 꽃수술, 꽃받침, 씨앗 등

#### 13.6.3. 예문

- <헛뿌리: PT\_PART>는 고착작용을 하지만 간혹 흡수작용도 한다.
- <씨: PT\_PART> 받아서 심으신다.
- 요즘처럼 <잎: PT\_PART>이 누렇게 물들어 떨어질 때면 황기를 수확하기에 적기다.
- 보리수 <열매: PT\_PART>에는 아미노산과 비타민C가 풍부해서 기관지에는 물론 피부 미용에도 좋다고 하는데요.

#### 13.6.4. 주의 사항

- (식물의 부분) PT\_PART는 눈으로 확인 가능한 식물의 부분만 대상으로 태깅한다.
- '꽃', '뿌리', '잎', '줄기', '수술', '씨앗' 등 식물의 특정 부분의 명칭을 PT\_PART로 태깅한다.
- (껍질) '껍질'은 식물에만 존재하는 것이 아닌 일반 물질이므로 PT\_PART로 태깅하지 않는다.
- (씨앗 이름) '해바라기씨', '목화씨' 는 PT\_PART로 태깅한다.

- 해바라기씨 → <해바라기씨: PT\_PART>
- 목화씨 → <목화씨: PT\_PART>
- (보통명사 꽃, 식물의 부분으로서의 꽃) 보통명사 '꽃'과 식물의 부분으로서의 '꽃' 둘 다 PT\_PART로 태깅한다.
  - 태권도의 꽃으로 불리는 최종량급 → 태권도의 <꽃: PT\_PART>으로 불리는 최종량급
  - 화분에 꽃이 탐스럽게 피어 있다. → 화분에 <꽃: PT\_PART>이 탐스럽게 피어 있다.

## 13.7. PT\_OTHERS

### 13.7.1. 정의 및 설명

- 위의 식물의 세부분류에 포함되지 않는 기타 식물 명칭

### 13.7.2. 분석 대상

- 이끼류
  - 이끼, 물이끼, 솔이끼, 뿔이끼 등
- 버섯류
  - 버섯, 송이버섯, 밤버섯, 싸리버섯, 송로, 표고버섯, 느타리버섯, 목이버섯, 양송이버섯, 영지버섯, 백화고 등
- 해조류
  - 김, 미역, 다시마, 파래, 청태, 매생이, 톳, 대황, 모자반, 청각, 우뚝가사리 등
- 그 외 기타 식물
  - 선인장, 콩나물 등

### 13.7.3. 예문

- <상황버섯: PT\_OTHERS>에 상어지느러미까지.
- 사실 갯지렁이는 모랫속에 들어 있는 미생물과 <해초: PT\_OTHERS>를 먹고 삽니다.
- 따라서 생선은 적당히 먹되 마늘이나 <미역: PT\_OTHERS> 등 수은해독에 효과 있는 음식을 함께 먹으라는 조언입니다.
- <선인장: PT\_OTHERS>에는 주로 가시가 있고 잎이 없다.
- 양상하게 변한 볼품없는 <선인장: PT\_OTHERS>이 제작진의 시야에 들어왔다.

### 13.7.4. 주의 사항

- (기타식물류) 콩나물은 PT\_OTHERS로 태깅한다.
  - 콩나물은 콩을 발아시켜 키운다. → <콩나물: PT\_OTHERS>은 콩을 발아시켜 키운다.
  - 이럴 때는 채소 중에서도 배추와 양배추, 마늘, 콩나물 같은 흰색 채소를 많이 드세요.

→ 이럴 때는 채소 중에서도 배추와 양배추, 마늘, <콩나물: PT\_OTHERS> 같은 흰색 채소를 많이 드세요.

## 14. MATERIAL

---

### 14.1. MT\_ELEMENT

#### 14.1.1. 정의 및 설명

- 원소명
  - 원소기호와 번호를 할당받은 주기율표에 등록된 물질

#### 14.1.2. 분석 대상

- 원소 물질
  - 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨

#### 14.1.3. 예문

- 원소 물질
  - 30일 KDDI와 배터리 납품업체 교세라는 일부 모델에 채용한 <리튬: MT\_ELEMENT> 이온 충전지에서 화재 및 파열사고가 발생하면서 이들 중 판매된 물량을 회수하기로 했다고 블룸버그통신이 보도했다.
  - 서울우유가 국내 치즈 중 <나트륨: MT\_ELEMENT> 함유량이 가장 낮은 '유기농 양팡 치즈' 제품 4종을 새롭게 선보인다고 22일 밝혔다.
  - 한국에너지기술연구원(원장 황주호 · 에너지연)은 기존에 난방이나 급탕용으로만 이용하던 태양열을 활용해 미래의 에너지원이라 불리는 <수소: MT\_ELEMENT>를 생산하는 기술을 개발했다고 17일 밝혔다.
  - 여기에 지난해 가격을 동결했던 일부 산업용가스메이커도 올 들어 산업용 전기료 인상을 명분 삼아 <산소: MT\_ELEMENT>, <질소: MT\_ELEMENT>, <아르곤: MT\_ELEMENT> 등 산업용가스가격을 소폭 올리고 있다.
  - 호남석유화학은 30일 오후 3시에 <탄소: MT\_ELEMENT>복합재 전문기업인 데크항공㈜

의 지분 인수를 위한 주식매매계약을 테크항공의 대주주인 ㈜테크와 체결했다.

#### 14.1.4. 주의 사항

- (원소기호) 해당 원소기호가 문맥상 원소를 가리키는 것이 분명할 때에만 MT\_ELEMENT로 태깅한다.
  - NA 함량이 높은 음식 → <NA: MT\_ELEMENT> 함량이 높은 음식
  - NA 대통령 트럼프 → <NA: LCP\_COUNTRY> <대통령: CV\_POSITION> <트럼프: PS\_NAME>
    - 이 때의 NA는 나트륨의 원소기호로 쓰인 것이 아닌 NorthAmerica의 약자이므로 MT로 태깅하지 않는다.
- (복합어 내부 태깅) 원소명 MT는 해당 원소명 및 금속물의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다.
  - 산소마스크 → <산소: MT\_ELEMENT>마스크

### 14.2. MT\_METAL

#### 14.2.1. 정의 및 설명

- 금속물

#### 14.2.2. 분석 대상

- 금, 은, 동, 칼슘, 황금, 황산철, 알루미늄, 아연, 구리, 수은, 아말감

#### 14.2.3. 예문

- NYMEX에서 <구리: MT\_METAL> 12월물 가격은 파운드당 3.45센트 내린 2.9320달러에 거래를 마쳤다.
- 6세기 말경에는 <금: MT\_METAL>의 광석에서 <금: MT\_METAL>을 추출하는 데에 이미 <수은: MT\_METAL>이 이용되었다.
- <칼슘: MT\_METAL>의 흡수도 방해할 뿐만 아니라 우리 몸에 들어와서 신장결석도 유발할 수 있으니까
- <황금: MT\_METAL> 같은 버들 사이로 왕래한다.
- <알루미늄: MT\_METAL>이라는 이름은 백반(白礬)에서 유래한다

#### 14.2.4. 주의 사항

- (복합어 내부 태깅) 금속물은 복합어 내부에서도 본래의 금속물적인 의미를 유지하고 있다면 MT\_METAL로 태깅한다.

- 마그네슘리본 → <마그네슘: MT\_METAL>리본
- 칼슘왕 미꾸라지 → <칼슘: MT\_ELEMENT>왕 미꾸라지
- (유일태깅 원칙) 금속물질 중에서 금속 원소는 MT\_METAL과 MT\_ELEMENT가 모두 태깅 가능한데, MT\_METAL로 태깅한다.
  - 리튬, 알루미늄, 구리, 철, 아연, 은, 금, 납 등
  - 티타늄 재질이라 가볍다. → <티타늄: MT\_METAL> 재질이라 가볍다.
- (준금속 원소) 준금속 원소는 MT\_ELEMENT로 태깅한다.
  - 붕소, 규소, 게르마늄(저마늄), 비소, 안티모니, 텔루륨, 폴로늄
  - 게르마늄으로 키운 농산물 → <게르마늄: MT\_ELEMENT>으로 키운 농산물

### 14.3. MT\_ROCK

#### 14.3.1. 정의 및 설명

- 암석 명칭
- 보석 명칭

#### 14.3.2. 분석 대상

- 암석 명칭
  - 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암
- 보석 명칭
  - 사파이어, 다이아몬드, 루비, 진주, 크리스탈

#### 14.3.3. 예문

- 파란색 <사파이어: MT\_ROCK>와 <다이아몬드: MT\_ROCK>로 된 약혼반지
- 특히 지경부는 연탄제조용 중간재인 <석탄: MT\_ROCK>가격은 생산원가 증가분만큼 가격을 인상했다.
- 이와 관련해 <연탄: MT\_ROCK> 공장도가를 개당 373.5원으로 동결하고, <연탄: MT\_ROCK>원료 <무연탄: MT\_ROCK> 가격은 생산원가 인상을 감안해 15% 인상했다.
- <옥: MT\_ROCK>과 같은 고가의 재료도 사용된 만큼 기린상은 중국인들에게 인기가 높다.
- 구매거부 中파워에 <철광석: MT\_ROCK> <유연탄: MT\_ROCK> 가격 2년 만에 하락
- 실제로 인도산 <철광석: MT\_ROCK>의 중국향 현물가격(스팟가격)은 지난 6월 t당 124달러에서 8월 149달러로 20%이상 상승했고 최근 오퍼가격 역시 155달러까지 올랐다.
- 반짝이는 <크리스탈: MT\_ROCK>은 체코를 찾는 관광객들에게 아름다운 기억을 남게 하는 프라하의 명물이지만 가짜 <크리스탈: MT\_ROCK>이 쏟아져 나오면서 그 명성이 흔들리고 있습니다.



- 산책로에는 목재 데크와 <황토: MT\_ROCK> 포장을 사용해 쉽게 다닐 수 있도록 하고 주변 경관과 어울릴 수 있도록 할 계획입니다.

#### 14.3.4. 주의 사항

- (암석명칭+-류) 암석 명칭의 경우 -류 등과 결합하여 암석 종류를 나타낼 때 암석만을 태깅 대상으로 한다.
  - 화강암류 → <MT\_ROCK: 화강암>류
- (광석류) MT\_METAL에 해당하는 물질의 광석의 경우, 복합어 내부의 해당 물질만 태깅하면 전체의 의미를 나타내지 못하게 되므로 최장 단위 태깅 원칙을 적용하여 MT\_ROCK으로 태깅한다.
  - 철광석, 금광석, 은광석, 동광석 등
  - 거대 금광석 발견 → 거대 <금광석: MT\_ROCK> 발견

### 14.4. MT\_CHEMICAL

#### 14.4.1. 정의 및 설명

- 화학물질 명칭
  - 화학의 대상이 되는 물질
  - 화학 작용을 일으키는 물질
  - 화학 산업으로 만들어진 인공의 물질
  - 영양분이 있는 물질
  - 생명체의 기관이나 조직을 촉진, 억제하는 호르몬 물질
- 화학물질의 하위 유형
  - 고체 화학물질
  - 액체 화학물질
  - 기체 화학물질
  - 영양소 및 호르몬

#### 14.4.2. 분석 대상

- 고체 화학물질
  - 파라핀, 레이온, 나프탈렌, 수산화나트륨, 염화나트륨, 탄산나트륨 등
- 액체 화학물질
  - 경유, 휘발유, 석유, 플루오르화수소, 액화석유가스, 과산화수소, 에탄올, 니트로벤젠, 리놀레산 등
- 기체 화학물질

- 암모니아, 황화수소, 이산화탄소, 이산화황, 일산화탄소, 일산화질소, 네온 등
- 영양성분 및 호르몬
  - 헤모글로빈, 프리온, 홀로수린, 키틴질, 콜레스테롤, 지방, 에스트로겐, 사포닌, 비타민 등

#### 14.4.3. 예문

- 고체 화학물질
  - <탄산나트륨: MT\_CHEMICAL>을 이용하여 제조한 순환잔골재의 품질 특성
  - <나프탈렌: MT\_CHEMICAL>을 대체할 천연 탈취제를 찾고 계시나요?
  - <수산화나트륨: MT\_CHEMICAL>과 <탄산칼륨: MT\_CHEMICAL> 수용액을 전기분해한 알칼리 이온수
- 액체 화학물질
  - 한국석유공사 등에 따르면 국제 유가 하락 등에 힘입어 21일 0시를 기해 제주지역 기름 값이 <휘발유: MT\_CHEMICAL>는 30원, <경유: MT\_CHEMICAL>는 40원 하락했다.
  - 이처럼 보통 살균제로 쓰이는 <과산화수소: MT\_CHEMICAL>가 만병통치약이라는 주장이 퍼지고 있습니다.
  - 공업용 <에탄올: MT\_CHEMICAL>로 '손소독제' 만든 일당 적발
- 기체 화학물질
  - 화장실에서 <암모니아: MT\_CHEMICAL> 냄새가 나는 원인은?
  - 배달대행업체 '전기이륜차' 도입 <이산화탄소: MT\_CHEMICAL> 배출 줄인다.
  - <CO2: MT\_CHEMICAL>를 줄이는 탄소포인트제예요.
- 영양성분 및 호르몬
  - 보리수 열매에는 <아미노산: MT\_CHEMICAL>과 <비타민C: MT\_CHEMICAL>가 풍부해서 기관지에는 물론 피부미용에도 좋다고 하는데요.
  - 우리 몸은 <콜레스테롤: MT\_CHEMICAL>이 축적되는 걸 아주 싫어한대요.
  - 또한 <오메가3지방산: MT\_CHEMICAL>이 풍부해 고지혈증과 치매 예방 등에도 효과적이라고 합니다.

#### 14.4.4. 주의 사항

- (유일태깅 원칙) 화학물질 중에 원소에 해당하는 경우에는 MT\_ELEMENT와 MT\_CHEMICAL이 모두 가능한데, 원소명에 해당하는 경우에는 무조건 MT\_ELEMENT로 태깅한다.
  - 산소생산 해양 박테리아 → <산소: MT\_ELEMENT>생산 해양 박테리아
- (영양성분 부류) "미네랄"은 '생체의 생리 기능에 필요한 광물성 영양소'로 칼륨, 나트륨, 칼슘, 인, 철 따위 모두 포함하는 상위 용어지만 MT\_CHEMICAL의 하위부류 중 '영양성분'에 포함되므로 MT로 태깅한다. "지방", "단백질"도 이와 비슷한 경우이다.
- (추출물 부류) "석류추출물"과 같은 추출물 부류는 그 추출물 전체를 물질의 이름으로 보기 어려우므로 전체를 MT로 태깅하지 않는다. "사과향", "오렌지향"과 같은 향의 부류도

이와 비슷한 경우이다.

- 구 전체를 태깅하지 않지만, 과일명은 PT\_FRUIT로 태깅한다.
  - 치자 추출물 → <치자: PT\_FRUIT> 추출물
  - 망고 추출물 → <망고: PT\_FRUIT> 추출물
  - 사과 향 → <사과: PT\_FRUIT> 향
  - 오렌지 향 → <오렌지: PT\_FRUIT> 향
- (영양제 부류) 화학물질 하위 부류 중 영양성분 및 호르몬은 영양제 TMM\_DRUG로도 태깅 가능한데, 문맥과 적정 술어를 통해 그 혼동을 해소한다.
  - 비타민이 부족해 면역력이 떨어진다. → <비타민: MT\_CHEMICAL>이 부족해 면역력이 떨어진다.
    - ~이 부족하다, ~이 풍부하다, ~을 섭취하다 등
  - 어제 비타민을 복용했다. → 어제 <비타민: TMM\_DRUG>을 복용했다.
    - ~을 복용하다, ~을 처방하다, ~을 먹다 등
- (식물 화학물질) 식물의 어떤 부분이나 물질, 또는 그러한 명칭에 대해서 MT\_CHEMICAL로 태깅하는 것이 적절한지 혼동되는 경우가 있다.
  - 옷, 송진, 꽃가루, 창포물 등
    - "옷"은 옷나무의 진을 이르기도 하는데, 색소로 쓰이기도, 피부병을 유발하기도 한다. 이러한 물질의 이름을 "옷"으로 명명했다고 보기는 어려우므로 MT\_CHEMICAL로 태깅하지 않는다. "송진", "꽃가루", "창포물"도 이와 비슷한 경우이다.
  - 사포닌, 폴리페놀, 피톤치드, 모노페놀, 안토시아닌 등
    - "사포닌"은 '식물계에 널리 존재하는 배당체'로 사전에 실려있는 화학물질로, MT\_CHEMICAL로 태깅한다. "폴리페놀", "피톤치드", "모노페놀", "안토시아닌"도 이와 비슷한 경우이다.

## 15. TERM

---

이 부류는 위에서 정의된 세부 개체명 이외의 개체명 용어를 포함한다. 아래에 정의된 세부 개체명만을 대상으로 하고, 그 이외의 개체명은 제외한다.

### 15.1. TM\_COLOR

#### 15.1.1. 정의 및 설명

- 색 및 빛 이름
- 색의 하위 유형
  - 계통색 이름
  - 관용색 이름
  - 빛 이름

### 15.1.2. 분석 대상

- 계통색 이름
  - 갈색, 갈회색, 갈흑색, 검정색, 남색, 노란색, 녹색, 녹회색, 보라색, 분홍색, 빨간색, 연노란색, 연두색, 연분홍색, 자주색, 적자색, 주황색, 진분홍색, 청록색, 초록색, 파란색, 황갈색, 회갈색, 회색, 회적색, 회청색, 흰색
- 관용색 이름
  - 계란색, 국방색, 당근색, 대추색, 딸기색, 루비색, 물색, 박하색, 벚꽃색, 벽돌색, 복숭아색, 사과색, 산호색, 살구색, 석류색, 수박색, 숙색, 연지색, 우유색, 장미색, 코코아색, 토마토색, 풀색, 하늘색, 호두색, 호박색, 홍차색
- 빛 이름
  - 갈색빛, 구릿빛, 개나리빛, 검은빛, 노란빛, 녹빛, 다홍빛, 보랏빛, 은빛, 핏빛, 흰빛
- 톤 이름
  - 화이트 톤, 블루 톤, 핑크 톤, 레드 톤, 그레이 톤, 그린 톤

### 15.1.3. 예문

- 계통색 이름
  - 여자 화장실은 <빨간색: TM\_COLOR>, 남자 화장실은 <파란색: TM\_COLOR>으로 표시하고 있다.
  - 경기도 남양주시에서 <흰색: TM\_COLOR> 강아지를 발견하여 보호하고 있다.
  - <노란색: TM\_COLOR>, <주황색: TM\_COLOR> 등 알록달록한 색상의 비즈로 만든 액세서리가 사람들의 관심을 끌고 있다.
- 관용색 이름
  - <국방색: TM\_COLOR> 옷을 입은 남자가 걸어가고 있다.
  - 화장품에서는 <복숭아색: TM\_COLOR>을 잘 살린 색조 상품들이 돋보인다.
  - 올해는 <대추색: TM\_COLOR>이나 <도라지색: TM\_COLOR> 등 은은하면서 자연스러운 색상이 인기를 끌고 있다.
- 빛 이름
  - 자작나무는 <흰빛: TM\_COLOR>을 띠고 있는 나무껍질을 가지고 있다.
  - 선미는 신곡과 같은 <보랏빛: TM\_COLOR> 아이 메이크업을 완벽하게 소화했다.
  - 29일부터 제주에서는 가을 감성을 담은 <감빛: TM\_COLOR> 옷 패션쇼를 시작한다.

- 톤 이름
  - 식빵언니 김연경의 집은 <화이트 톤: TM\_COLOR>의 깔끔한 인테리어가 돋보였다.
  - 디자이너들은 올봄 의상에 다양한 <블루 톤: TM\_COLOR>을 사용했다.
  - 그녀는 <레드 톤: TM\_COLOR>의 드레스를 입고 뛰어난 미모를 뽐냈다.

#### 15.1.4. 주의 사항

- (태깅 범위) 색깔을 표현하는 형용사와 결합된 경우만 전체를 TM\_COLOR로 태깅한다.
  - 이 꽃은 검은 자주색이다. → 이 꽃은 <검은 자주색: TM\_COLOR>이다.
  - 저기 진한 갈색 옷을 입은 사람이 내 남자친구야. → 저기 진한 <갈색: TM\_COLOR> 옷을 입은 사람이 내 남자친구야.
- (특수문자) 기호가 삽입되어 있는 경우에는 기호를 포함하여 태깅한다.
  - 올 가을은 팔죽+포도주스색이 인기를 끌고 있습니다. → 올 가을은 <팔죽+포도주스색: TM\_COLOR>이 인기를 끌고 있습니다.
- (마르살라) 사전적 의미가 색이 아니거나 '-색, -톤'과 함께 쓰이지 않아도 문맥상 '색'을 의미할 경우 TM\_COLOR로 태깅한다.
  - 올 가을에는 적갈색 '마르살라'가 패션 업계를 물들이고 있다. → 올 가을에는 <적갈색: TM\_COLOR> '마르살라: TM\_COLOR'가 패션 업계를 물들이고 있다.
- (임의로 붙인 색상명) 화장품 제조사에서 특정 색상에 임의로 붙인 이름은 TM\_COLOR로 태깅하지 않는다.
- 입생로랑 워터틴트 '바그드 루쥬'와 '인베리답'은 말을 더해 줘도 부족한 예쁜 컬러입니다. → 입생로랑 워터틴트 '<바그드 루쥬: AFW\_OTHER\_PRODUCTS>'와 '<인베리답: AFW\_OTHRE\_PRODUCTS>'은 말을 더해줘도 부족한 예쁜 컬러입니다.
  - '바그드 루쥬, 인베리답, 햇살 한 조각, 어제보다 로제, 로맨스엔 로즈' 등 화장품 제조사에서 임의로 붙인 이름은 AFW\_OTHER\_PRODUCTS로 태깅한다.
  - 제가 가진 색상은 TRD01 레드 TPK04 맘모 푸시아 색상이에요. → 제가 가진 색상은 TRD01 <레드: TM\_COLOR> TPK04 맘모 푸시아 색상이에요.
- 'TRD01'과 같이 임의로 붙인 색상명 또한 TM\_COLOR로 태깅하지 않는다.
  - 오렌즈의 '스칸디 헤이즐'은 블랙핑크 지수와 제니가 화보 속에서 착용한 렌즈이다. → 오렌즈의 '<스칸디 헤이즐: AFW\_OTHER\_PRODUCTS>'은 블랙핑크 지수와 제니가 화보 속에서 착용한 렌즈이다.
    - 화장품 외에도 렌즈, 휴대폰 등 제조사에서 임의로 붙인 색상명은 TM\_COLOR로 태깅하지 않는다.
- (흰검, 검파 등) '흰검, 검파' 등과 같이 색상을 줄여 나타낸 경우 전체를 TM\_COLOR로 태깅한다.
  - 흰금 vs 파흑 색깔 논란, 보는 것이 믿는 것이다? → <흰금: TM\_COLOR> vs <파흑: TM\_COLOR> 논란, 보는 것이 믿는 것이다?

- 나이키 코르테즈 흰검 여자 운동화로 딱! → 나이키 코르테즈 <흰검: TM\_COLOR> 여자 운동화로 딱!

## 15.2. TM\_DIRECTION

### 15.2.1. 정의 및 설명

- 방향
  - 방향의 하위 유형
    - 방위
    - 기타 분류

### 15.2.2. 분석 대상

- 방향
  - 상측, 하측, 상향, 서북향, 아래쪽, 앞, 앞뒤, 앞쪽, 옆쪽, 오른쪽, 오른쪽, 왼쪽, 왼쪽, 우측, 위쪽, 동남향, 뒤쪽, 남향, 좌측, 하향
- 방위
  - 동, 서, 남, 북, 중앙, 북서, 북동, 남서, 남동, 동서, 동북
- 기타 분류
  - 지역의 부분: 서부, 남부, 동부, 북부, 중부, 최남단, 최동단, 최북단, 최서단
  - 위치: 위, 아래, 옆

### 15.2.3. 예문

- 방향
  - 학교를 기준으로 <오른쪽: TM\_DIRECTION>에는 서점이, <왼쪽: TM\_DIRECTION>에는 은행이 있습니다.
  - 최근에는 <동남향: TM\_DIRECTION> 또는 <서남향: TM\_DIRECTION> 집을 선호한다.
  - 롯데 자이언츠 한동희는 <우측: TM\_DIRECTION> 어깨 통증을 호소했다.
  - 승관은 <좌측: TM\_DIRECTION> 발목 인대 파열로 수술하였고, 세븐틴은 당분간 12인 체제로 활동한다.
  - 방위
    - 남극이 <서쪽: TM\_DIRECTION>부터 녹는 이유는 무엇일까?
    - 여기서 <동쪽: TM\_DIRECTION>으로 1km만 가 보세요.
- 지역의 부분
  - 오늘 <중부: TM\_DIRECTION> 지역은 강한 비가 쏟아질 것으로 보입니다.
  - 강원 <북부: TM\_DIRECTION> 양돈 농가를 대상으로 출입 통제를 실시하고 있습니다.

- 대한민국 <최남단: TM\_DIRECTION>은 어디일까?
- <최북단: TM\_DIRECTION> 캠핑장으로 놀러 오세요!
- 위치
  - 책상 <위: TM\_DIRECTION>에는 책이 있고, 책상 <아래: TM\_DIRECTION>에는 가방이 놓여 있다.

#### 15.2.4. 주의 사항

- (방향) '좌우, '앞뒤'는 TM\_DIRECTION으로 태깅한다.
  - 고개를 좌우로 돌려 보세요. → 고개를 <좌우: TM\_DIRECTION>로 돌려 보세요.
  - 파도가 심해 배가 앞뒤로 기우뚱거린다. → 파도가 심해 배가 <앞뒤: TM\_DIRECTION>로 기우뚱거린다.
- (태깅 대상) 명사, 명사구만을 대상으로 한다.
  - 남쪽으로 물러났던 태풍이 북상했다. → <남쪽: TM\_DIRECTION>으로 물러났던 태풍이 북상했다.
    - '북상'은 이동의 의미를 지니며, 명사 및 명사구를 대상으로 하기에 '북상하다'는 태깅하지 않는다.
- (남부권) '남부권'과 같이 '영역, 범위'를 의미하는 경우 TM\_DIRECTION으로 태깅하지 않는다.
  - 국토교통부는 경기 북부에 주택 33만호를 공급할 예정이며, 경기 북부권 공공택지 입주자 모집은 올해 시작된다. → 국토교통부는 경기 <북부: TM\_DIRECTION>에 주택 33만호를 공급할 예정이며, 경기 북부권 공공택지 입주자 모집은 올해 시작된다.
    - '서부산권'과 같이 방향과 지명이 함께 나타나는 경우, 사전에 등재되어 있지 않기 때문에 '서부산'만 TM\_DIRECTION으로 태깅하지 않는다.

### 15.3. TM\_CLIMATE

#### 15.3.1. 정의 및 설명

- 기후 분류 명칭

#### 15.3.2. 분석 대상

- 기후 분류 명칭
  - 열대기후, 건조기후, 온대기후, 냉대기후, 한대기후, 열대우림기후, 사바나기후, 스텝기후, 사막기후, 온난습윤기후, 온난동계건조기후, 지중해성기후, 냉대습윤기후, 아열대기후, 고산기후, 도시기후, 동안기후, 서안기후, 대륙성 기후, 해양 기후, 해양성 기후, 사막기후, 스텝기후

### 15.3.3. 예문

- 최근 <건조기후: TMI\_CLIMATE>로 인해 화재 사고가 많으니 유의하시기 바랍니다.
- W.P.코펜은 이것을 <한대기후: TMI\_CLIMATE>와 <냉대기후: TMI\_CLIMATE>의 경계로 하였다.
- 거제는 위도 34-35도에 자리하고, 연평균 강우량이 1,860mm로 전형적인 <해양성 기후: TMI\_CLIMATE>에 속한다.
- <동안기후: TMI\_CLIMATE>란 대륙 동안에 나타나는 <대륙성기후: TMI\_CLIMATE>를 말하는 것으로 우리나라가 이에 해당한다.
- <서안기후: TMI\_CLIMATE>는 연중 바다의 영향을 받기 때문에 항상 온난습윤한 기후 특성을 가진다.

### 15.3.4. 주의 사항

- (태깅 대상) 기후 명칭만을 태깅한다.
  - 동안기후는 서안기후에 비하여 여름에는 고온다습하고, 겨울에는 한랭건조한 것이 특징이다. → <동안기후: TMI\_CLIMATE>는 <서안기후: TMI\_CLIMATE>에 비하여 여름에는 고온다습하고, 겨울에는 한랭건조한 것이 특징이다.
    - '고온다습, 한랭건조'와 같은 기후 특성과 관련된 명칭은 태깅하지 않는다.
  - 열대는 <열대기후: TMI\_CLIMATE> 지역으로서 고온다습하고 고무, 코코야자, 코코아 등의 플랜테이션이 영위되고 있다.
    - '열대, 아열대, 온대' 등이 문맥상 '지리, 지대'를 의미할 때는 태깅하지 않는다.

## 15.4. TM\_SHAPE

### 15.4.1. 정의 및 설명

- 모양
- 모양의 하위 유형
  - 형태
  - 무늬

### 15.4.2. 분석 대상

- 형태
  - 유선형, 일자형, 원뿔꼴, 원뿔형, 일자, 원, 삼각형, 사각형, 직사각형, 팔각형, 오각형, 피라미드형, 방울 패턴, 기하학적 패턴, 타원, 동그라미, 동그러미, 둥근형, 선형, 소용돌이, 수상형, 실꼴, 십자꼴, 십자형, 이삭꼴, 피라미드형, 달걀모양, 둥근모양, 뚜껑모양, 타원모



- 양, 사다리꼴모양, H형, 일자형, 원통형, 삼각뿔 모양, 사각뿔 모양
- 무늬
  - 가로무늬, 구름무늬, 줄무늬, 점무늬

### 15.4.3. 예문

- 형태
  - <삼각형: TM\_SHAPE>의 종류에는 <정삼각형: TM\_SHAPE>, <이등변삼각형: TM\_SHAPE>, <직각삼각형: TM\_SHAPE> 등이 있다.
  - 최근에는 <사각형: TM\_SHAPE> 수박도 인기를 끌고 있다.
  - 발톱은 <일자 모양: TM\_SHAPE>으로 깎아야 한다.
  - <둥근 모양: TM\_SHAPE>의 감자처럼 생긴 과자가 출시되었다.
  - <삼각뿔 모양: TM\_SHAPE> 강남 아파트 분양받아 볼까?
- 무늬
  - 복원된 세종의 자격루 표면에서는 그동안 보이지 않았던 <구름 무늬: TM\_SHAPE>가 선명하게 나타났다.
  - 산 정상에 올라가면 밑에 있는 사람들이 작은 <점무늬: TM\_SHAPE>처럼 보인다.

### 15.4.4. 주의 사항

- (태깅 범위) 'N-형/모양/패턴'의 경우 전체를 TM\_SHAPE로 태깅한다.
  - 컴퓨터 시뮬레이션을 통해 설계한 피라미드형 콘서트홀은 음악사상 유래가 없는 것이다. → 컴퓨터 시뮬레이션을 통해 설계한 <피라미드형: TM\_SHAPE> 콘서트홀은 음악사상 유래가 없는 것이다.
  - H형 책상을 무료나눔하니 필요한 분은 연락 주세요. → <H형: TM\_SHAPE> 책상을 무료나눔하니 필요한 분은 연락 주세요.
- (태깅 대상) '형/모양/패턴'이 생략되어도 문맥상 모양의 의미를 지니면 TM\_SHAPE로 태깅한다.
  - 달고나가 하트, 별 등 다양한 모양으로 되어 있어 보는 사람들의 눈길을 사로잡았다. → 달고나가 <하트: TM\_SHAPE>, <별: TM\_SHAPE> 등 다양한 모양으로 되어 있어 보는 사람들의 눈길을 사로잡았다.
- (무늬) 'OO무늬'까지만 TM\_SHAPE로 태깅하고, 실선, 점선, 직선, 곡선은 태깅하지 않는다.

## 15.5. TM\_CELL\_TISSUE\_ORGAN

### 15.5.1. 정의 및 설명

- 세포의 명칭

- (인체 내의) 장기, 기관 및 조직 명칭

### 15.5.2. 분석 대상

- 세포 명칭
  - 극세포, 간상세포, 딸세포, 신경세포, 정원세포, 정모세포, 가시세포, 각막세포, 그물세포, 주세포, 황색세포, 유전자, 아교세포
- (인체 내의) 장기, 기관 및 조직 명칭
  - 근육, 혈액, 침샘, 뉴런, 망치뼈, 모루뼈, 등자뼈, 갈비뼈, 늑연골, 반지연골, 방패연골, 쇄기연골, 복사뼈, 턱관절, 인대, 뼈, 고막, 식도, 골반, 고환, 동맥, 정맥, 모세혈관, 내장, 콩팥, 비장, 신장

### 15.5.3. 예문

- 세포
  - <원추세포: TM\_CELL\_TISSUE\_ORGAN>와 섞여서 같은 열에 들어서 있으나, <간상세포: TM\_CELL\_TISSUE\_ORGAN>의 수가 훨씬 더 많다.
  - 분열 전의 <모세포: TM\_CELL\_TISSUE\_ORGAN>(<母細胞: TM\_CELL\_TISSUE\_ORGAN>)에 대해서 <딸세포: TM\_CELL\_TISSUE\_ORGAN>라고 하며 <낭세포: TM\_CELL\_TISSUE\_ORGAN>(<娘細胞: TM\_CELL\_TISSUE\_ORGAN>)라고도 한다.
  - <신경세포: TM\_CELL\_TISSUE\_ORGAN>는 <뉴런: TM\_CELL\_TISSUE\_ORGAN> 전반에 영양을 공급해 주며, 돌기의 수에 따라 무극(apolar) · 단극(unipolar) · 쌍극(bipolar) · 다극(multipolar) 등으로 분류된다.
  - <표피: TM\_CELL\_TISSUE\_ORGAN>의 주요 기능은 식물체의 보호이다.
  - <정원세포: TM\_CELL\_TISSUE\_ORGAN>는 유사분열을 거듭해서 일정한 수준까지 증식한 다음 성장과 성숙을 거쳐서 <정모세포: TM\_CELL\_TISSUE\_ORGAN>(<精母細胞: TM\_CELL\_TISSUE\_ORGAN>)가 된다.
- (인체 내의) 장기, 기관 및 조직 명칭
  - 운동 후 마시는 프로틴이 맛도 좋고 <근육: TM\_CELL\_TISSUE\_ORGAN>에도 좋다.
  - 최근 오투기에서는 <침샘: TM\_CELL\_TISSUE\_ORGAN>을 자극하는 다양한 신제품을 출시했다.
  - <고막: TM\_CELL\_TISSUE\_ORGAN>에 도달한 음파는 <고막: TM\_CELL\_TISSUE\_ORGAN>을 진동시키는데 그 진동은 <망치뼈: TM\_CELL\_TISSUE\_ORGAN>·<모루뼈: TM\_CELL\_TISSUE\_ORGAN>·<등자뼈: TM\_CELL\_TISSUE\_ORGAN>와 <이소골: TM\_CELL\_TISSUE\_ORGAN>을 거쳐 <내이: TM\_CELL\_TISSUE\_ORGAN>에 전달된다.
  - <식도: TM\_CELL\_TISSUE\_ORGAN>의 내강(內腔)은 엄지손가락 굵기 정도이지만, 음식물이 통과할 때는 상당히 확장된다.
  - <골반: TM\_CELL\_TISSUE\_ORGAN>은 그 속에 성기 ·분비기관 ·소화기관의 하부를 수용하

고 보호하는 동시에 양쪽 다리와 연결되어 몸을 떠받치고 있다.

- 고단백을 장기간 섭취할 경우 <콩팥: TM\_CELL\_TISSUE\_ORGAN>기능이 감소할 수 있다.
- <비장: TM\_CELL\_TISSUE\_ORGAN>은 <혈액: TM\_CELL\_TISSUE\_ORGAN>을 걸러주는 필터로서의 역할을 한다.

#### 15.5.4. 주의 사항

- (태깅 범위) 눈으로 보이지 않는 부분만 TM\_CELL\_TISSUE\_ORGAN으로 태깅한다.
  - 현재 방영하고 있는 '사이코지만 괜찮아'에는 안구 정화 비주얼 커플이 등장한다. → 현재 방영하고 있는 '사이코지만 괜찮아'에는 <안구: AM\_PART> 정화 비주얼 커플이 등장한다.
    - '안구'는 일반적으로 눈으로 확인할 수 있는 부분이기에 AM\_PART로 태깅한다.
- (혈관) '동맥', '정맥', '모세혈관'은 태깅 대상이나, '혈관'은 태깅하지 않는다.
  - 혈관은 동맥, 모세혈관, 정맥을 총칭한다. → 혈관은 <동맥: TM\_CELL\_TISSUE\_ORGAN>, <모세혈관: TM\_CELL\_TISSUE\_ORGAN>, <정맥: TM\_CELL\_TISSUE\_ORGAN>을 총칭한다.
- (인공관절) '인공관절'도 TM\_CELL\_TISSUE\_ORGAN으로 태깅한다.
  - 관절의 노화나 손상 정도가 심각하면 인공관절 수술을 고려해야 한다. → <관절: TM\_CELL\_TISSUE\_ORGAN>의 노화나 손상 정도가 심각하면 <인공관절: TM\_CELL\_TISSUE\_ORGAN> 수술을 고려해야 한다.
- (복합어) '소 간, 손가락 뼈'는 띄어쓰기와 상관없이 사전에 등재되어 있기에 통합하여 태깅한다.
  - 소 간은 눈물이 많은 반려견에게 효과적이다. → <소 간: TM\_CELL\_TISSUE\_ORGAN>은 눈물이 많은 반려견에게 효과적이다.
  - 손가락 뼈에서 소리가 나는 이유는 무엇일까? → <손가락 뼈: TM\_CELL\_TISSUE\_ORGAN>에서 소리가 나는 이유는 무엇일까?
- (비뇨기) 기관을 총칭하는 '비뇨기'는 TM\_CELL\_TISSUE\_ORGAN으로 태깅하지 않는다.
  - 오늘부터 콩팥, 방광 등 하복부·비뇨기 초음파도 건강보험이 적용됩니다. → 오늘부터 <콩팥: TM\_CELL\_TISSUE\_ORGAN>, <방광: TM\_CELL\_TISSUE\_ORGAN> 등 하복부·비뇨기 초음파도 건강보험이 적용됩니다.
- (피부, 살, 이, 혀) '세포, 조직, 기관' 등을 의미할 때만 TM\_CELL\_TISSUE\_ORGAN으로 태깅하며, '피부, 살, 이, 혀, 털' 등은 AM으로 태깅한다.
  - 충치를 예방하려면 3분 안에 꼭 이를 닦아야 한다. → 충치를 예방하려면 3분 안에 꼭 <이: AM\_PART>를 닦아야 한다.
  - 잘못된 화장 습관은 속눈썹 빠짐의 원인이므로 주의해야 한다. → 잘못된 화장 습관은 <속눈썹: AM\_PART> 빠짐의 원인이므로 주의해야 한다.

## 15.6. TMM\_DISEASE

### 15.6.1. 정의 및 설명

- 증상 및 질환

### 15.6.2. 분석 대상

- 증상
  - 기침, 가래, 경련, 딸국질, 발열, 재채기, 변비, 저체온증, 편두통, 멍, 출혈
- 질환
  - 내과질환: 역류성식도염, 직장암, 폐결핵, 각기병, 위장염, 하혈, 늑막염, 패혈증, 식중독
  - 외과질환: 콜린성두드러기, 헤르페스, 기면증, 뇌경색, 뇌출혈, 뇌진탕, 치매, 찰과상, 골절
  - 산부인과 및 소아과질환: 구루병, 소아마비
  - 안과 및 치과질환: 안검염, 결막염, 콩다래끼, 녹내장
  - 이비인후과질환: 이하선염, 인후염, 고막염, 인후암, 후두염
  - 기타질환: 페스트, 스페인독감

### 15.6.3. 예문

- 증상
  - 나이가 들면 <기침: TMM\_DISEASE>을 하는 것이 당연하다고 생각하면 안 된다.
  - 거짓말을 하면 <딸국질: TMM\_DISEASE>을 하게 된다.
  - 이 질병의 증상은 <발열: TMM\_DISEASE>과 <재채기: TMM\_DISEASE>이다.
  - 자고 일어났더니 온몸에 <멍: TMM\_DISEASE>이 들었다.
- 질환
  - 이처럼 <당뇨병: TMM\_DISEASE>은 꾸준한 관리가 필요한 질환이다.
  - 유재석은 <고소공포증: TMM\_DISEASE>을 이겨내고 번지점프대에서 뛰어내렸다.
  - 위검사 결과 <위염: TMM\_DISEASE> 증세가 있었다.
  - 연구팀은 지난해 <자궁경부암: TMM\_DISEASE>으로 진단된 환자의 난소를 떼어내 섭씨 영하 196도에 냉장보관했다.
  - <냉방병: TMM\_DISEASE>은 <감기: TMM\_DISEASE> 수준으로 대처하는 경우가 많으나, <폐렴: TMM\_DISEASE> 증상까지 발생할 수 있어 주의가 필요하다.
  - 목욕만 잘 시켜도 강아지 <피부병: TMM\_DISEASE>을 예방할 수 있다.
  - 습한 장마철에는 <눈병: TMM\_DISEASE>에 걸리기 쉽다.

### 15.6.4. 주의 사항

- (바이러스) 질병의 원인을 의미하는 '바이러스'는 TMM\_DISEASE로 태깅한다.
- 어린 아이들은 바이러스에 감염되기 쉽기에 특별히 조심해야 한다. → 어린 아이들은 <바

이러스: TMM\_DISEASE>에 감염되기 쉽기에 특별히 조심해야 한다.

- 문맥을 고려하여 '-이/가 죽다, -이/가 번식하다, -이/가 자라다, -이/가 성장하다'처럼 미생물을 의미할 경우에는 AM\_OTHERS로, '-에 걸리다, -에 감염되다, -에 시달리다, -을/를 앓다, -을/를 예방하다, -을/를 치료하다, -을/를 고치다, -이/가 낫다'처럼 질병의 원인을 의미할 경우에는 TMM\_DISEASE로 태깅한다.
- (세균) '세균'은 질병의 원인으로 볼 수 있으나 AM\_OTHERS로 태깅한다.
  - 뇌수막염은 세균이 원인이기에 항생제가 필수이다. → <뇌수막염: TMM\_DISEASE>은 <세균: AM\_OTHERS>이 원인이기에 항생제가 필수이다.
- (태깅 범위) 질환 전체를 TMM\_DISEASE로 태깅한다.
  - 급성 폐쇄각 녹내장은 빠른 시간 내에 안압의 상승으로 발생하는 질환으로, 즉시 치료를 받아야 한다. → <급성 폐쇄각 녹내장: TMM\_DISEASE>은 빠른 시간 내에 안압의 상승으로 발생하는 질환으로, 즉시 치료를 받아야 한다.
- (태깅 대상) 의학적으로 치료가 필요한 증상, 질환만을 태깅한다.
- 월요병에 걸린 직장인들이 지친 몸을 이끌고 출근한다. → 월요병에 걸린 직장인들이 지친 몸을 이끌고 출근한다.
  - '중2병, 월요병, 연예인병, 관심병, 가난병' 등 이름만 병인 것들은 태깅하지 않는다.
- (-질환) '체장 질환' 등 'N-질환'은 전체를 TMM\_DISEASE로 태깅한다.
  - 체중이 감소할 경우 체장 질환을 염두하고 검사를 받아야 한다. → 체중이 감소할 경우 <체장 질환: TMM\_DISEASE>을 염두하고 검사를 받아야 한다.
- (강박) '강박' 등과 같은 정신 질환도 TMM\_DISEASE로 태깅한다.
  - 현대인의 절반 이상이 우울증, 강박 등 정신 질환을 앓고 있는 것으로 조사되었다. → 현대인의 절반 이상이 <우울증: TMM\_DISEASE>, <강박: TMM\_DISEASE> 등 <정신 질환: TMM\_DISEASE>을 앓고 있는 것으로 조사되었다.
- (후유증) '후유증'은 TMM\_DISEASE로 태깅한다.
  - 김장하다보면 온몸이 육신육신하는 후유증이 생긴다. → 김장하다보면 온몸이 육신육신하는 <후유증: TMM\_DISEASE>이 생긴다.
- (잡티, 기미) '잡티, 기미'도 TMM\_DISEASE로 태깅한다.
  - 요즘 고생이 많은지 여자의 얼굴에 기미가 끼었다. → 요즘 고생이 많은지 여자의 얼굴에 <기미: TMM\_DISEASE>가 끼었다.

## 15.7. TMM\_DRUG

### 15.7.1. 정의 및 설명

- 약 및 의약품명
- 약의 하위 유형

- 외용약
- 내복약
- 기타

### 15.7.2. 분석 대상

- 외용약
  - 물파스, 파스, 고약, 살균제, 연고
- 내복약
- 헤파린, 아스피린, 졸피뎀, 멜라토닌, 항생제, 항히스타민제, 아세트아미노펜, 히알루론산, 소염제, 페니실린, 항우울제, 제산제, 소화제, 진정제, 수면제, 구충제, 근이완제, 영양제, 변비약(완화제), 각성제, 건강제, 건강보조제
- 기타
  - 모기향, 생맥산

### 15.7.3. 예문

- 또한 전신성 혈관내 응고(disseminated intravascular coagulation)와 혈전생성을 억제하며 투여 후 출혈빈도는 <헤파린: TMM\_DRUG>보다 적게 나타난다.
- 고래에서는 인슐린을 비롯하여 고기·심장·간에서 채취하는 엑스류(類) 외에, 건조갑 상선·경랍(鯨蠟)·취장건조분말·젤라틴·<헤파린: TMM\_DRUG> 등을 얻을 수 있다.
- <근이완제: TMM\_DRUG>가 이러한 근수축을 방해하는 작용을 한다.
- <근이완제: TMM\_DRUG>는 크게 두 가지 종류로 구분되는데, 중추신경에 작용하는 것과 말초신경에 작용하는 것이 있다.
- 백화(白花) 계통의 제충국(除蟲菊)이 해충구제에 효과가 있으며, 그 분말을 점착제로 고결시켜서 성형한 것이 <모기향: TMM\_DRUG>이다.
- <소화제: TMM\_DRUG>에는 위장운동을 항진시켜 소화액의 분비를 높이는 약제와 소화액의 분비부족을 보충하는 소화효소제가 있다.
- <아스피린: TMM\_DRUG>은 최초로 합성된 해열, 소염 진통제이다.
- <구충제: TMM\_DRUG> 복용 후의 <완하제: TMM\_DRUG>(<緩下劑: TMM\_DRUG>)는 충분한 양을 제때에 복용한다.

### 15.7.4. 주의 사항

- (태깅 대상) 의약품명만 TMM\_DRUG로, 구체적인 상품명은 AFW\_OTHER\_PRODUCTS로 태깅한다.
  - 약국이 허락한 유일한 마약은 어린이 영양제 텐텐이다. → 약국이 허락한 유일한 마약은 어린이 <영양제: TMM\_DRUG> <텐텐: AFW\_OTHER\_PRODUCTS>이다.
  - 작은 상처에는 후시딘연고와 같이 쉽게 찾아볼 수 있는 연고를 바르면 된다. → 작은 상

처에는 <후시딘연고: AFW\_OTHER\_PRODUCTS>와 같이 쉽게 찾아볼 수 있는 <연고: TMM\_DRUG>를 바르면 된다.

- (한방 처방) 한방에서의 처방도 TMM\_DRUG로 태깅한다.
  - 더운 여름에 기운을 보강하기 위한 처방으로는 생맥산이 대표적이다. → 더운 여름에 기운을 보강하기 위한 처방으로는 <생맥산: TMM\_DRUG>이 대표적이다.
    - '-짓다|조제하다|처방하다' 등과 같이 적정 술어가 사용되었을 때만 TMM\_DRUG로 태깅한다.
- (탈모 예방 영양제) '탈모 예방 영양제'와 같은 경우에는 TMM\_DISEASE, TMM\_DRUG를 분리하여 태깅한다.
  - 젊었을 때부터 탈모 예방 영양제를 먹어야 한다. → 젊었을 때부터 <탈모: TMM\_DISEASE> 예방 <영양제: TMM\_DRUG>를 먹어야 한다.

## 15.8. TMI\_HW

### 15.8.1. 정의 및 설명

- 하드웨어 용어 및 전자제품
- 이 외의 장치명, 장치를 구성하는 부품 등을 포함

### 15.8.2. 분석 대상

- 하드웨어 용어
  - 입력장치: 키보드, 마우스, 스캐너, 마이크, 조이스틱
  - 중앙처리장치: 마이크로 프로세서, 그래픽 프로세서, 사운드 프로세서, 메인 보드, 그래픽 카드, 사운드 카드, 랜 카드
  - 기억장치: 임시기억장치(램, 시스템 메모리, 캐쉬 메모리), 영구기억장치(롬), 플로피 디스크
  - 저장장치: 하드디스크 드라이브
  - 출력장치: 모니터, 프린터, 스피커
- 전자제품: 컴퓨터, 모니터, 팩스, 카메라, 디카, 디지털카메라, 냉장고, 전자레인지, 핸드폰, 마우스, 전화(기), 노트북, 계산기, 라디오, 에어컨, 세탁기, 텔레비전, 히터기, 네비게이션, 방수 카메라, 충전기, 스마트폰, 사운드바, 디지털 오디오, 엘피, 수화기, 카세트, 게임기
- 장치: 망원경, 인공위성, 이동통신단말기, 무선화상수신부, 회동지지바, 연동지지바, 광섬유 격자, 광전송선, 광써큘레이터, 광검출부, 충격측정부
  - 로봇: 뷰리오시티, 오퍼튜니티(화성 탐사 로봇), 스피릿(화성 탐사 로봇)
- 장치를 구성하는 부품: 소자 소켓, 감각 센서, 후각 센서, 감지 센서
- 코드: 바코드, QR코드

### 15.8.3. 예문

- 하드웨어 용어
  - <하드디스크: TMI\_HW>는 자성체를 입힌 원판형 알루미늄 기판을 회전시키면서 자료를 저장하고 읽어 내도록 한 보조기억장치의 하나로서
  - <마이크로프로세서: TMI\_HW>는 주기억장치에 저장되어 있는 명령어를 인출하여 해독하고, 해독된 명령어를 실행하며 실행 결과를 다시 주기억장치에 저장할 수 있는 기능 등을 자동으로 수행함과 동시에 입출력 장치들과도 데이터 교환을 수행한다.
  - <롬: TMI\_HW>(<ROM: TMI\_HW>)은 read only memory의 약칭이다. 컴퓨터의 판독전용 기억장치를 말한다.
  - <디스켓: TMI\_HW>은 기록밀도에 따라 2D·2DD·2HD 등으로 분리하고, 크기에 따라 5.25인치 · 3.5인치 등으로 분리하고 있다.
  - <자기테이프: TMI\_HW>는 한때 컴퓨터 보조기억장치로 널리 쓰였으나 자료 처리가 순차적으로만 이루어져 처리 시간이 느리고, 순차적으로 구성되어 있지 않은 자료를 처리하는 데는 실용적이지 못하다는 단점이 있다.
  - <사운드카드: TMI\_HW>는 4스피커 지원 여부, 3D사운드 지원 여부 리버브나 코러스 같은 이펙트 지원, 풀 듀플렉스 등의 기능에 따라 차이가 난다.
- 전자제품
  - <김치냉장고: TMI\_HW> 용량이 1백50ℓ인 모델은 쌀을 27kg까지 담을 수 있고 1백82ℓ짜리는 39kg을 보관할 수 있다.
  - 지난 8월 LG전자 창원공장에서 <냉장고: TMI\_HW>·<세탁기: TMI\_HW>를 판매하면서 연예인을 초청, 오락 프로그램을 곁들여 5시간 동안 생방송했다.
  - 올 들어 지속된 경기침체로 <에어컨: TMI\_HW> 예약 판매량이 30% 이상 줄어 고민하던 가전업체들은 예상보다 빠른 <에어컨: TMI\_HW> 특수를 맞아 생산라인을 풀가동하기 시작했다.
- 장치명
  - 발사된 <인공위성: TMI\_HW> 가운데 가장 많은 비율을 차지하는 것은 과학위성이다.
  - 임흥빈 삼성증권 테크팀장은 "<이동통신단말기: TMI\_HW>의 갑작스런 출하 지연과 신규모델로 인한 판매관리비 증가도 영업이익률 하회에 기여했다"고 밝혔다.
- 장치를 구성하는 부품
  - <감지 센서: TMI\_HW>를 통해 각각의 냄새에 대한 반응을 전송한다.
- 코드
  - 계산대의 여종업원은 너무나 지루하다는 듯이 물건들의 <바코드: TMI\_HW>를 기계에 읽힌 다음 비닐봉지에 담아 주었다.

### 15.8.4. 주의 사항



- (태깅 대상) 정확한 상품명으로 나온 경우는 AFW\_OTHER\_PRODUCTS로 태깅하며, 그 외 보통명사에 해당하는 제품들은 TMI\_HW로 태깅한다.
- 스마트폰에서는 아이폰3GS보다 갤럭시 노트20이 인기를 끌고 있다. → <스마트폰: TMI\_HW>에서는 <아이폰3GS: AFW\_OTHER\_PRODUCTS>보다 <갤럭시 노트20: AFW\_OTHER\_PRODUCTS>이 인기를 끌고 있다.
- (전자제품) '카메라'와 같은 제품 유형은 TMI\_HW로, 세부 제품명은 TMI\_MODEL로 태깅한다.
  - 초소형 슬라이드 카메라폰 E800이 사람들의 주목을 받고 있다. → <초소형 슬라이드 카메라폰: TMI\_HW> <E800: TMI\_MODEL>이 사람들의 주목을 받고 있다.
- (장치명) 특허 문서에는 특성상 장치명, 모듈명 등이 존재하는데, 개체명 대상이 아니지만 특허 분야에서는 중요한 키워드가 되기에 '-장치, -모듈, -부' 등도 태깅한다.
  - 검출기란 정량적인 값으로 변환하고 이해하기 쉬운 형태로 표시하는 장비로, 광 검출기, 입자 검출기 등이 있다. → 검출기란 정량적인 값으로 변환하고 이해하기 쉬운 형태로 표시하는 장비로, <광 검출기: TMI\_HW>, <입자 검출기: TMI\_HW> 등이 있다.
    - '검출기, 단말기'와 같이 단어 자체만으로 특정한 개체를 인지하기 어려운 경우는 태깅하지 않는다.
  - 스마트폰 충전기 앞으론 하나로/모든 기종 표준형으로 통일. → <스마트폰: TMI\_HW> <충전기: TMI\_HW> 앞으론 하나로/모든 기종 표준형으로 통일.
    - 충전기와 같이 일반적으로 개체가 무엇을 하는 것인지 인지가 가능한 경우는 태깅한다.
- (LED) 'LED'는 장치를 구성하는 부품으로 보고 TMI\_HW로 태깅한다.
  - 삼성전자는 액정표시장치의 일종인 LED 출시를 공식화했다. → 삼성전자는 액정표시장치의 일종인 <LED: TMI\_HW> 출시를 공식화했다.

## 15.9. TMI\_SW

### 15.9.1. 정의 및 설명

- 소프트웨어 용어
- 소프트웨어 하위 유형
  - 시스템 소프트웨어
  - 응용 소프트웨어
  - 기타 분류

### 15.9.2. 분석 대상

- 시스템 소프트웨어

- 운영 체제: 윈도우XP, 윈도우10, 리눅스, 도스
- 언어 번역 프로그램: 어셈블러, 컴파일러, 인터프리터
- 유틸리티 프로그램: 알집, 알약, 안랩, 네이버 백신, 고클린, 반디집, 알송
- 응용 소프트웨어
  - 한컴오피스 한글, MS워드, 파워포인트, 엑셀, 곰플레이어, CAD, 그림판, 인터넷 뽐, 익스플로러, 크롬, 오토캐드, 일러스트레이터, 워드 뷰어
- 기타 분류
  - 악성 코드: 트로이목마, 웜 바이러스, 님다바이러스
  - 프로그래밍 언어: 자바, 파이썬, C언어, 코볼

### 15.9.3. 예문

- 시스템 소프트웨어
  - 경찰청은 31일 "지난 17일 미국 버지니아의 한 인터넷 서비스 제공업체에 접속한 해커가 강릉시의 한 PC방 <리눅스: TMI\_SW> 서버의 운영권을 획득하고, 이를 거점으로 전국 250여 대 컴퓨터(170여개 기관)에 침투를 시도했다"고 밝혔다.
  - 회사 컴퓨터 <알집: TMI\_SW>을 업데이트하면서 악성 프로그램을 내려받게 되었다.
- 응용 소프트웨어
  - 대학생들의 대다수는 <MS워드: TMI\_SW>보다 <한글: TMI\_SW>을 많이 사용하는 것으로 조사되었다.
  - <곰플레이어: TMI\_SW>로 유명한 미디어 전문 IT기업 곰앤컴퍼니(대표 이병기)가 자사 동영상 편집 프로그램을 새로 출시했다고 15일 밝혔다.
- 악성 코드
  - <님다 바이러스: TMI\_SW>가 다른 바이러스보다 더 빠른 속도와 경로를 통해 확산되고 있다.
  - <웜 바이러스: TMI\_SW>는 시스템 파일을 제어하고 시스템 숨김 파일이 보이지 않도록 설정한다.
- 프로그래밍 언어
  - <C 언어: TMI\_SW>는 프로그램 오류를 쉽게 발견하기 위한 기능은 부족하지만, 고수준 언어에서 자주 볼 수 있는 기술상의 제약이 적기 때문에 오히려 프로그래밍하기 쉬운 편리한 언어로 평가되며, 프로그램을 간결하게 쓰기 위하여 많은 연구를 한 언어이다.
  - 최근에는 <파이썬: TMI\_SW> 등 컴퓨터 프로그래밍 언어를 배울 수 있는 학원이 증가하고 있다.

### 15.9.4. 주의 사항

- (유일 태깅 원칙) AF\_WARES와 중복되는 경우 TMI\_SW를 우선적으로 태깅한다.
- (태깅 대상) TMI\_SW는 '운영체제'와 같은 제품 유형을, TMI\_MODEL은 '안드로이드'와 같은

세부 제품명을 나타낸다.

- (일반 인터넷 용어) '사이트, 홈페이지, 링크' 등과 같은 일반 인터넷 용어는 태깅하지 않는다.
  - 익스플로러에서는 안전한 사이트가 맞으면 주소창이 연두색으로 물들고 자물쇠 모양이 뜬다. → <익스플로러: TMI\_SW>에서는 안전한 사이트가 맞으면 주소창이 연두색으로 물들고 자물쇠 모양이 뜬다.
- (시스템 및 프로그램) 특허 문서에는 특성상 '-시스템, -프로그램' 등이 존재하는데, 개체명 대상이 아니나 중요한 키워드기에 TMI\_SW로 태깅한다.
  - 송은숙 스마트비투엠 부사장(사진)은 “지난 1월부터 중소기업에 적합한 명함관리 프로그램을 개발했다”며 “버려지는 명함을 회사 자산으로 만들어 줄 다양한 카테고리를 탑재한 것이 특징”이라고 말했다. → 송은숙 스마트비투엠 부사장(사진)은 “지난 1월부터 중소기업에 적합한 <명함관리 프로그램: TMI\_SW>을 개발했다”며 “버려지는 명함을 회사 자산으로 만들어 줄 다양한 카테고리를 탑재한 것이 특징”이라고 말했다.
    - '데몬, 명함 관리 시스템, 위성위치확인시스템, 지리정보시스템' 등도 TMI\_SW로 태깅한다.

## 15.10. TMI\_SITE

### 15.10.1. 정의 및 설명

- URL 주소
- URL의 하위 유형
  - http
  - mailto
  - ftp

### 15.10.2. 분석 대상

- http: <http://www.tving.com>, <http://www.juso.go.kr>, <http://jtbc.joins.com>
- https: <https://www.naver.com>, <https://www.korean.go.kr>, <https://www.daum.net>
- mailto: <mailto:apple@test.com>
- ftp: <ftp://ftp.microsoft.com>

### 15.10.3. 예문

- http
  - 반려동물을 맡겨둘 수 있는 영업장은 동물보호관리시스템(<[www.animal.go.kr](http://www.animal.go.kr): TMI\_SITE>)에서 확인하시기를 바랍니다.

- 문화재청은 국가문화유산포털(<http://www.heritage.go.kr: TMI\_SITE>)의 문화재 정보 접근성과 검색방식 등을 개선하여 이달부터 시행한다.
- https
  - 국립국어원(<https://www.korean.go.kr: TMI\_SITE>) 온라인가나다를 통해 어문 규범이나 어법 등을 문의할 수 있다.
- mailto
  - '<mailto:admin@test.it: TMI\_SITE>'와 같이 국세청을 사칭한 메일주소를 주의하시기 바랍니다.
- ftp
  - 마이크로소프트 사의 FTP 주소는 '<ftp://ftp.microsoft.com: TMI\_SITE>'입니다.

#### 15.10.4. 주의 사항

- (URL 생략) 'http, www' 등 URL의 일부가 생략되어도 TMI\_SITE로 태깅한다.
  - 코로나 최신 기사는 news.kbs.co.kr에서 확인하실 수 있습니다. → 코로나 최신 기사는 <news.kbs.co.kr: TMI\_SITE>에서 확인하실 수 있습니다.

### 15.11. TMI\_EMAIL

#### 15.11.1. 정의 및 설명

- 이메일 주소

#### 15.11.2. 분석 대상

- 이메일 주소: test@naver.com, test@hanmail.net, test@etri.re.kr

#### 15.11.3. 예문

- 참가하실 분은 이메일(<test@naver.com: TMI\_EMAIL>)로 참가 신청서를 작성하여 보내주시기 바랍니다.

### 15.12. TMI\_MODEL

#### 15.12.1. 정의 및 설명

- 제품의 모델명

#### 15.12.2. 분석 대상

- 제품의 모델명: SCH-X100, SPH-X9750, FQ22L8DRA2, CAP-100SW

### 15.12.3. 예문

- 예약판매하는 그램 17 대표 모델(모델명: <17Z90N-VA50K: TMI\_MODEL>) 출하가는 214만원이다.
- 프로기아가 상급자용 드라이버 <TR500: TMI\_MODEL>을 출시했다.

### 15.12.4. 주의 사항

- (태깅 대상) TMI\_HW/SW와 혼동되는 경우가 많은데, 세부 모델명인 경우에만 TMI\_MODEL로 태깅한다.
  - 초소형 슬라이드 카메라폰 E800이 사람들의 주목을 받고 있다. → <초소형 슬라이드 카메라폰: TMI\_HW> <E800: TMI\_MODEL>이 사람들의 주목을 받고 있다.
    - TMI\_HW 는 '카메라'와 같은 제품 유형이고, TMI\_MODEL은 'D800'과 같은 세부 제품명을 나타낸다.
- (부품의 상품명) 부품의 상품명은 AFW\_OTHER\_PRODUCTS로 태깅하고 모델명은 TMI\_MODEL로 태깅한다.
  - 새로 출시되는 삼성 갤럭시에는 스냅드래곤801(APQ8074AA)의 탑재 가능성이 크다. → 새로 출시되는 삼성 갤럭시에는 <스냅드래곤801: AFW\_OTHER\_PRODUCTS>(<APQ8074AA: TMI\_MODEL>)의 탑재 가능성이 크다.

## 15.13. TMI\_SERVICE

### 15.13.1. 정의 및 설명

- IT 서비스 용어
- IT 서비스 하위 유형
  - 애플리케이션
  - 인터넷 사이트
  - 특정 서비스

### 15.13.2. 분석 대상

- 애플리케이션: 인스타그램, 트위터, 카카오톡, 리브메이트, 페이스북, 페이스북
- 인터넷 사이트명
  - 포털 사이트: 네이버, 다음, 구글, 야후, 라이코스, 네티앙, 아이팝콘
  - 웹사이트명: 네이버TV, 도그티비, 고양이라서 다행이야, 세모지, 수미사, 수제나라, 오늘의 유머, 올레닷컴, 아프리카 TV, 트위치

- 유튜브 채널명: 씨티비, 땅고부부, 딤디, 빼약스핏, 오딜
- 특정 서비스: 와이브로 서비스, DMB 서비스, 위성 DMB, 익스박스라이브, IPTV

### 15.13.3. 예문

- 애플리케이션
  - 소방청은 이번 달 22일부터 <카카오톡: TMI\_SERVICE>으로 응급환자와 상담하는 서비스를 시행한다.
  - 이문환 행장은 간편결제 플랫폼 <페이북: TMI\_SERVICE>을 이용한 제휴 서비스 출시를 준비하고 있다.
- 인터넷 사이트명
  - 대표적인 인터넷 포털사이트로는 <네이버: TMI\_SERVICE>, <다음: TMI\_SERVICE> 등이 있다.
  - 며칠 전 <고양이라서 다행이야: TMI\_SERVICE> 카페에 들어갔다가 재미있는 사진을 발견했다.
  - <땅고부부: TMI\_SERVICE>와 함께 재미있는 운동을 해 보세요.
- 특정 서비스
  - SK텔레콤은 2018년을 끝으로 <와이브로 서비스: TMI\_SERVICE>를 종료했다.
  - 스마트폰이 대중화됨에 따라 <위성DMB: TMI\_SERVICE>는 역사 속으로 사라졌다.

### 15.13.4. 주의 사항

- (특정 서비스) 익스박스라이브와 같이 서비스만을 대상으로 하며, 다른 세부유형과 혼동되는 경우에는 다른 세부유형에게 우선권을 부여한다.
- (아날로그 방송) '아날로그 방송, 디지털 방송'은 특허분야에서 특정한 서비스를 의미하는 확률이 높기에 TMI\_SERVICE로 태깅한다.
  - 하지만 시험방송 기간 중에는 일부 프로그램을 제외하고는 단순히 아날로그방송 프로그램을 디지털로 바꿔 내보내는 방식이어서 화질이 조금 떨어진다. → 하지만 시험방송 기간 중에는 일부 프로그램을 제외하고는 단순히 <아날로그방송: TMI\_SERVICE> 프로그램을 디지털로 바꿔 내보내는 방식이어서 화질이 조금 떨어진다.
- (인터넷 사이트) '팩스 서비스'와 같이 일반적인 서비스, '웹사이트, 홈페이지, 포털사이트' 등과 같은 일반적인 인터넷 용어는 태깅하지 않는다.
  - 범죄자 신상공개 웹사이트인 디지털교도소는 많은 사람들의 관심을 받고 있다. → 범죄자 신상공개 웹사이트인 <디지털교도소: TMI\_SERVICE>는 많은 사람들의 관심을 받고 있다.
- (애플리케이션) 온라인 서비스를 의미하는 일반적인 용어 'SNS'는 태깅하지 않는다.
  - SNS는 페이스북과 트위터와 같은 다양한 서비스를 말한다. → SNS는 <페이스북: TMI\_SERVICE>과 <트위터: TMI\_SERVICE>와 같은 다양한 서비스를 말한다.

- (복합어) '인형스타그램, 먹스타그램' 등과 같은 복합어 '-스타그램'은 태깅하지 않는다.
  - 최근 인스타그램에서는 조카스타그램, 일상스타그램 등 다양한 해시태그를 통해 게시물을 접할 수 있다. → 최근 <인스타그램: TMI\_SERVICE>에서는 조카스타그램, 일상스타그램 등 다양한 해시태그를 통해 게시물을 접할 수 있다.
    - '인형스타그램, 커피스타그램' 등은 태깅 대상이 아니므로, '인형, 커피'와 같이 복합어 내부 요소 또한 태깅하지 않는다.
- (줄임말) '카카오톡'의 줄임말인 '카톡'은 TMI\_SERVICE로 태깅한다.
  - 세금 상담은 카톡으로 하세요. → 세금 상담은 <카톡: TMI\_SERVICE>으로 하세요.

## 15.14. TMI\_PROJECT

### 15.14.1. 정의 및 설명

- 프로젝트 명칭
- 프로젝트 하위 유형
  - IT 프로젝트
  - 국가 프로젝트
  - 회사 프로젝트

### 15.14.2. 분석 대상

- IT 프로젝트
  - 국가기간전산망사업, 국가디비 사업, 머큐리 계획, 아폴로 계획, 제미니 계획
- 국가 프로젝트
  - 4대강 사업, 메콩강유역개발사업, 두뇌한국21, 금강 뱃길 조성사업, 비전 2030 문화시민 도시 서울
- 회사 프로젝트
  - 그린IT사업, 무료치과진료사업, 삼성 희망디딤돌, 영상 개발 프로젝트

### 15.14.3. 예문

- 문재인 대통령은 <4대강 사업: TMI\_PROJECT> 감사 필요성을 제기하고, 사업 전반의 과정을 밝히는 데 집중했다.
- <머큐리 계획: TMI\_PROJECT>이란 미국 최초의 유인위성 발사계획을 말한다.
- 교육부는 대학원생의 장학금과 인건비를 지원하는 <두뇌한국21: TMI\_PROJECT> 기본계획을 확정하였다.
- 삼성전자가 보호종료 청소년을 위한 자립 지원 프로그램 '<삼성 희망 디딤돌: TMI\_PROJECT>'을 전국적으로 확대한다고 밝혔다.

#### 15.14.4. 주의 사항

- (태깅 범위) IT 계열이 포함되지만, IT 외의 국가/회사 프로젝트에 모두 TMI\_PROJECT로 태깅한다.
- (우주관련 계획) '머큐리 계획'과 같이 우주관련 계획은 IT 관련 프로젝트라 보고 TMI\_PROJECT로 태깅한다.
- (태깅 대상) '편찬 사업'과 같이 '사업'을 의미할 때는 TMI\_PROJECT로 태깅한다.
- (공익캠페인) '자연 보호 캠페인'과 같은 '운동'은 TMI\_PROJECT로 태깅하지 않는다.

### 15.15. TMIG\_GENRE

#### 15.15.1. 정의 및 설명

- 게임 장르

#### 15.15.2. 분석 대상

- 플레이 방식이나 게임의 성격에 따른 분류
  - RPG, AOS, FPS, 롤플레이팅 게임, 어드벤처 게임, 아케이드 게임, 시뮬레이션 게임, 전략 게임, 스포츠 게임, 퍼즐 게임, 리듬 게임, 레이싱 게임, 공포 게임, 슈팅게임, 추리 게임, 액션 게임, 판타지 게임, 무협 게임, 방치형 게임
- 플레이 환경이나 사용 기기에 따른 분류
  - PC게임, VR게임, 컴퓨터 게임, 온라인 게임, 모바일 게임, 콘솔 게임, 비디오 게임, 플래시 게임, 멀티 플랫폼 게임, 클라우드 게임
- 기타 분류
  - 사행성 게임, 인디 게임, 성인 게임, 가차 게임, 글로벌 게임, 한국사 컴퓨터 게임

#### 15.15.3. 예문

- 이미 일본과 세계 각지에서 인기를 끌고 있는 'Wii'는 이미 국내에서 140만 대 이상이 팔린 '닌텐도 DS Lite'의 인기를 힘입어 국내 <콘솔 게임: TMIG\_GENRE> 시장의 변화를 예고하고 있다.
- 한국의 <온라인 게임: TMIG\_GENRE>은 세계 시장에서도 큰 경쟁력을 갖추고 있다.
- 지금의 넥슨을 만든 일등공신 바람의 나라가 <모바일 게임: TMIG\_GENRE>으로 돌아왔다.
- 5G 환경이 갖추어짐과 함께 <클라우드 게임: TMIG\_GENRE>의 시대가 열리고 있다.

#### 15.15.4. 주의 사항

- (태깅 대상) 게임 장르상 분류가 아닌 단순 게임명은 TMIG\_GENRE로 태깅하지 않는다.



- 한국에 본격적인 게임 붐을 불러일으킨 RTS 스타크래프트는 '한국의 민속놀이'라는 우스갯소리도 듣는다. → 한국에 본격적인 게임 붐을 불러일으킨 <RTS: TMIG\_GENRE> 스타크래프트는 '한국의 민속놀이'라는 우스갯소리도 듣는다.

## 15.16. TM\_SPORTS

### 15.16.1. 정의 및 설명

- 레저:스포츠 기술 및 규칙 용어

### 15.16.2. 분석 대상

- 레저:스포츠 기술
  - 축구 기술 용어: 발리슛, 라보나킥, 세트피스, 오버헤드킥, 인프런트킥, 토킥, 간접프리킥, 스루패스, 크로스 트래핑, 드리블, 사포, 스로인, 티키타카, 도움주기
  - 농구 기술 용어: 버저비터, 레이업슛, 리바운드, 차징, 박스 아웃, 엘리웁, 크로스오버, 백다운, 스텝백
  - 야구 기술 용어: 스트라이크, 병살, 더블플레이, 안타, 홈런, 견제구, 고의사구, 번트, 벤치 클리어링, 보살, 더블헤더, 사구, 사사구, 희생플라이, 노히트노런, 배트플립, 투런포, 1점 홈런, 만루홈런
  - 기타 스포츠 용어
- 레저:스포츠 규칙
  - 축구 규칙 용어: 오프사이드, 캐링, 핸들링
  - 농구 규칙 용어: 실린더 룰, 라인크로스, 더블 드리블
  - 야구 규칙 용어: 보크, 몰수게임, 콜드게임
  - 기타 스포츠 용어

### 15.16.3. 예문

- 레저:스포츠 기술
  - 호날두는 환상적인 <오버헤드킥: TM\_SPORTS>으로 득점을 성공시켰다.
  - 경기 종료 직전 터진 <버저비터: TM\_SPORTS>는 정말 환상적이었다.
  - <희생플라이: TM\_SPORTS>로 1점 만회에 성공했다.
  - 1~1.5m 내외의 <퍼팅: TM\_SPORTS>을 수도 없이 놓쳤다.
- 스포츠 규칙
  - 부심은 기를 들어 보이며 <오프사이드: TM\_SPORTS>를 선언했다.
  - 손바닥을 위로 향해 공을 다루는 행위도 <더블드리블: TM\_SPORTS> 반칙에 해당한다.

#### 15.16.4. 주의 사항

- (태깅 범위) 방향성 정보 + 땅볼/안타/파울/홈런 등은 묶어서 TM\_SPORTS로 태깅한다.
  - 박용택은 이천웅의 1루 강습 안타로 2점을 선취했다. → 박용택은 이천웅의 <1루 강습 안타: TM\_SPORTS>로 2점을 선취했다.
- (스포츠 관련 용어) '볼넷, 삼진'과 같은 스포츠 규칙 관련 용어나 수량의 의미가 없는 1, 2루와 1, 2레인은 QT\_SPORTS가 아닌 TM\_SPORTS로 태깅한다.
  - 기아 김민식은 안타를 치고 1루로 향했다. → 기아 김민식은 안타를 치고 <1루: TM\_SPORTS>로 향했다.
  - 심규범은 <1이닝: QT\_SPORTS> <1볼넷: QT\_SPORTS> 무실점을 기록했다.
    - '1이닝', '1볼넷'은 QT\_SPORTS로 태깅한다.



## 참고 문헌

- 고석주(2007), 어휘의미망과 사전의 뜻풀이, 한국어의미학 24, 한국어의미학회.
- 김중환 · 최윤수 · 박태호(2015), 개체명 부착 말뭉치에서 자동 오류 수정, 한국 컴퓨터 종합 학술대회 논문집.
- 박동호(2000), 대상부류 개념의 응용언어학적 활용, 프랑스어문교육 10, 한국프랑스어문교육 학회.
- 박동호(2003), 의미부류 체계의 구축과 적용, 語學研究 39, 서울대학교 어학연구소.
- 박만규(2002), 대상부류 이론에서의 다의성, 프랑스어문교육 13, 한국프랑스어문교육학회.
- 박성재 · 차정원(2017), CNN을 이용한 대화와 같은 짧은 문장에서 개체명 인식, 한국 컴퓨터 종합학술대회.
- 박재연(2008), 고유명의 변별적 속성에 대하여, 한국어의미학 27, 한국어의미학회.
- 박재연(2012), 고유명 의미의 지시와 내포, 한국어의미학 37, 한국어의미학회.
- 박태호 · 신창욱 · 박성재 · 박다솔 · 차정원(2016), Rough Set을 이용한 형태소 품사 태깅 코 퍼스 오류 정량화, 한국 컴퓨터 종합학술대회 논문집.
- 박태호 · 차정원(2016), 커널 Ripple-Down Rule을 이용한 태깅 말뭉치 오류 자동 수정, 한국 정보과학회 논문지.
- 오진영, 차정원(2020), 뉴럴-심볼릭 구조 기반의 관계 추출, 한글 및 한국어정보처리 학술대 회.
- 이동혁(2007), 의미 범주 체계의 구축과 사전에서의 활용, 한국어의미학 24, 한국어의미학회.
- 이동혁(2010), 사전과 어휘의미망에서의 의미 기술, 한국사전학회 학술대회 발표논문집 8, 한 국사전학회.
- 이병모(2001), 명사의 하위 분류에 대하여, 국어학 64, 국어학회.
- 이성현(2007), 세종 전자 사전의 어휘 의미 부류 체계, 새국어생활 17, 국립국어원.
- 이은경(2016), 정신 공간 이론과 고유명사, 한국어의미학 51, 한국어의미학회.
- 이흥식(2016), 고유명사의 의미와 지시에 대하여, 한국학연구, 인하대학교 한국학연구소.
- 임홍빈 · 한재영(1993), 국어 어휘의 분류 목록에 대한 연구, 국립국어연구원.
- 차정원(2019), 개체명 분석 말뭉치 구축 연구보고서, 국립국어원.
- 최경봉(2001), 지식기반 구축을 위한 어휘의 의미 분류, 담화와인지 8, 담화 · 인지언어학회.
- 최윤수 · 차정원(2016), Word Embedding 자질을 이용한 한국어 개체명 인식 및 분류, 한국정 보과학회 논문지.
- 홍재성(2002), (21세기 세종계획)전자사전 개발분과 연구보고서, 문화관광부.
- Erik F. Tjong Kim Sang · Fien De Meulder(2003), Introduction to the CoNLL-2003 Shared Task:Language-Independent Named Entity Recognition, Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003.

<Abstract>

## 2020 Named Entity Corpus Research Analysis

With the advent of the 4th Industrial Revolution around the world and digital technologies such as Internet technology, artificial intelligence, and big data, Internet media has become the most important language communication environment. With social media and communities' development, more and more Internet media are used to communicate one's thoughts and experiences and communicate with others. In other words, the amount of language on the Internet becomes vast every day, and the value of the work of analyzing and converting it into data becomes increasingly important.

Reflecting the trend of these times, the '2020 Named Entity Corpus Research Analysis' project was conducted to build a large-scale, high-quality named entity analysis web corpus that can be used to develop artificial intelligence language processing technology and Korean language research in preparation for the 4th industrial revolution. We will also reorganize and refine the '19 National Institute of the Korean Language's Named Entity Annotation Guideline' created based on the 146 sub-classification systems of 'Named Entity Tag Set and Tagging Corpus' (Korea Information and Communication Technology Association, 2015) and 'Named Entity Guideline 2018' (Korea Electronics and Telecommunications Research Institute Language Intelligence Research Group, SW Contents Research Center, 2018. 12.) to be a standard for the construction of a named entity analysis corpus in the future.

Data.ly, the supervising project, and Hallym University, a collaborative project organization, received the "5 million words of primitive web corpus established by the National Institute of the Korean Language in 19," collected from blogs and web bulletin boards from the National Institute of the Korean Language, and worked on it for seven months. Data.ly was in charge of building a named entity, developing the system and tools, and

managing and training the building workers. Hallym University is to correct and supplement the existing guidelines and verify the corpus. I was in charge of building the evaluation corpus. To bring out successful results in an efficient manner, we have carried out this project as follows.

First, we used the named entity construction process and tools possessed by Data.ly.

We use a rough set (a method of calculating the classification boundary area by comparing the properties constituting each item in a set) and an automated system that automatically corrects errors using RDR (Ripple-Down Rule). This method specifies each case in which an exception condition occurs when there is one general rule that determines the classification. This system reduces errors by selecting items that are likely to be errors among the excepting conditions and correcting the sentences corresponding to the conditions. After applying the 146 named entity category to the automation system, we proceeded with the establishment, purification, and error detection of the entity name corpus. Since the automated system's work process is carried out collectively under the management of an expert from error detection to correction, it is a way to obtain both speed and accuracy of work. We compared the result calculated through this work with the evaluation corpus constructed by experts, calculated the result value, and corrected the error. Also, we reuse all the results constructed by the operator as data to improve the automation system's performance. We used a method of automatically generating an error correction rule using the error-correction pair that occurred during the operation, applying it to the entire corpus in the system, performing batch correction, and finally improving the corpus' quality.

Second, we used 'Kane,' a specialized tool for building a named entity that supports the crowd environment.

By allowing multiple users to work simultaneously in a crowd environment and immediately deliver the built data, we were able to proceed quickly. Also, a multi-faceted evaluation by many experts was performed to improve the quality of the resulting corpus and to make up for the shortcomings of quantitative evaluation by qualitative review by experts.

We have also improved work efficiency and reduced simple mistakes by using “Kane,” a program possessed by Data.ly, as a tool to build a corpus. Kane is a tool specialized for constructing a named entity, such as uploading, downloading corpus, searching word.

Lastly, many Korean linguistic experts from Hallym University and Kyunghee University participated.

Experts who have participated in related Korean language research and projects suggested problems and revised plans for existing guidelines and reflected opinions derived from discussions with the National Institute of the Korean Language. In addition, by constructing an evaluation corpus, the expert group’s results and the results of the working group could be compared and corrected, thereby reducing the error rate for the result and helping to increase the completeness of the corpus.

Since the Internet’s share of the language communication environment is increasing, web language materials are bound to increase in language research and language processing industries. However, compared to its high utilization, the amount of data for analyzing named entities in web languages is remarkably insufficient. We carried out this project to recognize the current situation and build a named entity analysis web corpus that can be used in various fields. The result of the ‘2020 named entity corpus research analysis’ project, which Data.ly and Hallym University participated in, will be used as public goods to be useful for Korean language research, artificial intelligence, and big data industries, and further, could be a stepping stone for the 4th industrial revolution.

**Keywords:** Web Corpus, Named Entity Corpus, intelligence language, Guideline, evaluation, construction

사업 책임자	차정원 (데이터리 대표)
사업 참여자	신서인 (한림대학교 교수)
	이선웅 (경희대학교 교수)
	오진영 (데이터리 이사)
	이수현 (한림대학교 시간강사)
	이한범 (한림대학교 시간강사)
	박진현 (한림대학교 박사수료)
	전상호 (한림대학교 박사과정)
	임윤정 (경희대학교 시간강사)
	황지영 (경희대학교 시간강사)
	박경희 (경희대학교 시간강사)
	박은실 (데이터리 전임연구원)
	배진영 (데이터리 전임연구원)
	최윤진 (데이터리 전임연구원)
담당 연구원	이승재(국립국어원 언어정보과장)
	이현주(국립국어원 언어정보과 학예연구관)
	오은비(국립국어원 언어정보과 연구원)

---

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9757

인쇄일: 2020년 12월 22일

발행일: 2020년 12월 22일

인 쇄: 서경기획

---

※ 이 책은 국립국어원의 용역비로 수행한 ‘2020년 개체명 말뭉치 연구 분석’ 사업의 결과물을 발간한 것입니다.





