

국립국어원 2021-01-16

발 간 등 록 번 호
11-1371028-000865-01

2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석

사업 책임자
차 정 원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업기간: 2021년 06월 ~ 2021년 12월

2021년 12월 17일

사업 책임자: 차 정 원 (데이터리)

사업 수행자

데이터리

한림대학교 산학협력단

사업 책임자

차정원

공동 연구원

신서인

사업 참여자

오진영, 이한범, 이수현, 박진현, 전상호, 김나현
김희인, 최양선, 남장현, 홍진기, 조아임, 오수진
임윤정, 황지영, 김희성, 김지영, 임고은, 이상금
박은실, 배진영, 최윤진, 서강원 외 32인

<사업 수행자>

데이터리 · 한림대학교 산학협력단

사업 책임자	차정원(데이터리 대표)
사업 참여자	신서인(한림대학교 교수)
	오진영(데이터리 이사)
	이한범(한림대학교 시간강사)
	이수현(한림대학교 시간강사)
	박진현(한림대학교 시간강사)
	전상호(한림대학교 시간강사)
	김나현(한림대학교 박사수료)
	김희인(한림대학교 연구원)
	최양선(한림대학교 석사)
	남장현(한림대학교 석사과정)
	홍진기(한림대학교 석사과정)
	조아임(경희대학교 박사수료)
	오수진(경희대학교 박사수료)
	임윤정(명지대학교 시간강사)
	황지영(경희대학교 석사)
	김희성(경희대학교 석사)
	김지영(경희대학교 석사과정)
	임고은(경희대학교 석사과정)
	이상금(한림대학교 시간강사)
	박은실(데이터리 전임연구원)
	배진영(데이터리 전임연구원)
	최윤진(데이터리 전임연구원)
	서강원(데이터리 전임연구원) 외 32인

2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석

오늘날 4차 산업혁명과 비대면 시대의 도래로 가장 주목받고 있는 분야는 빅데이터 및 인공지능 기술 개발이며, 이에 기반이 되는 대용량의 학습데이터 또한 수요가 계속해서 늘어나고 있다. 본 연구는 이러한 시대의 흐름을 파악하여 언어 데이터의 필요성을 인지하고 가장 활용도가 높은 개체명 분석과 개체 연결 분석을 진행하여 연구 및 사업 분야에 사용될 수 있는 말뭉치를 구축하였다. 또한 구축 작업에 필요한 지침을 명확하게 정립하여 이후에 이어질 연구의 기준과 방향성을 명확히 제시하는 것에 목적을 두었다.

‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업은 크게 개체명 분석 연구와 개체 연결 분석 연구로 나누어져 있다. 개체명 분석 과업은 다음과 같다. 첫째, 수집된 문어 및 구어 원시 말뭉치 300만 어절을 대상으로 개체명 분류 체계 (150개 세분류 기준) 내 해당 범주 표지를 부착하였다. 둘째, <2020년 개체명 분석 말뭉치 구축 지침>을 체계화하고 구체화하였다. 또한 선행 사업의 결과물인 ‘국립국어원 개체명 분석 말뭉치(버전1.0)’ 300만 어절의 개체명 분석 범위를 기존 부착되어 있던 15개 대분류에서 150개 세분류로 수정하는 작업도 함께 진행하였다.

개체 연결 분석 과업은 금년도에 신규로 추가되었다. 개체 연결 분석을 위한 방법론 수립과 말뭉치 구축 작업을 위한 지침을 작성하고, 개체명이 부착된 말뭉치 1,100만 어절을 대상으로 개체 연결 작업을 진행하였다.

마지막으로 개체명 및 개체 연결 말뭉치 1,100만 어절의 개체 정보를 추출해 개체명 사전을 구축하여 편리하게 데이터를 확인할 수 있도록 하였다. 개체명 사전은 개체명을 중심으로 정보가 통합된 구조이다. 따라서 동일한 의미의 개체명으로 그룹화되어 있어 중의성 해소에 용이하다. 또한 정보 검색 시스템, 질의 응답, 기계번역 등의 시스템 성능 향상에도 효과적으로 활용할 수 있다.

본 용역의 사업 기간은 약 7개월이며 수행 기관인 데이터리와 공동 수행 기관인 한림대학교는 국립국어원으로부터 대용량 원시 말뭉치를 전달받아 작업을

진행하였다. 전체적인 개체명 구축 프로세스 및 시스템, 도구 개발 그리고 구축 작업자 관리 및 교육 등의 총괄 책임과 관리는 데이터리에서 맡았다. 한림대학교는 국어학적 지식을 바탕으로 기존의 개체명 지침을 재정비하였다. 또한 개체 연결 연구 분석을 통해 개체 연결 구축 방법론을 정립하여 개체 연결 말뭉치 구축 작업을 진행함과 동시에 개체 연결 말뭉치 분석 지침을 수립하였다. 이와 관련한 구체적인 사업 수행 내용과 방법은 이후에 자세히 기술하도록 할 것이다.

본 사업을 통해 완성된 개체명 및 개체 연결 말뭉치와 구축 지침은 앞으로 국어 연구와 인공지능 및 빅데이터 산업에 도움이 될 수 있도록 공공재로 사용될 것이며, 개체명과 개체 연결 연구 분야의 발전에 기반이 되는 자료로 사용될 것이다.

주요어: 개체명, 개체 연결, 말뭉치, 지침, 구축

차 례

제1장 사업 개요

1. 사업 목적	3
2. 사업 범위	3
3. 사업 수행 기관	5
4. 사업 추진 경과	6

제2장 사업 수행 내용

1. 사업 수행 방안	9
1.1. 구축 방안	9
1.2. 지침 방안	10
1.3. 평가 방안	10
1.4. 관리 방안	10
2. 사업 수행 환경	11
2.1. 작업 공정	11
2.2. 구축 도구	14
2.3. 작업자 교육	22
2.4. 사업 수행 조직	27
3. 말뚝치 구축 및 납품	29
3.1. 말뚝치 구축	29
3.2. 말뚝치 납품	34
4. 검증 및 산출물 보고	35
4.1. 검증 절차	35
4.2. 검증 결과	36
4.3. 개체명 사전	36
4.4. 산출물	38
4.5. 사업 보고	44

차례

제3장 개체명 및 개체 연결 지침

1. 개체명 말뭉치 구축 지침 정비	49
1.1. 지침 주요 수정사항	49
1.2. 개체명 분석의 대원칙 정비	50
1.3. 세부 분류별 분석 지침 정비	51
2. 개체 연결 말뭉치 구축 지침 수립	56
2.1. 개체 연결 말뭉치 구축 지침 수립 과정	56
2.2. 개체명과 지식베이스(위키피디아) 문서 간의 관계 유형	57
2.3. 개체명과 지식베이스(위키피디아) 문서 연결의 원칙	60
2.4. 개체 연결 말뭉치 구축 지침 정비	66

제4장 결론

1. 사업 요약	73
2. 의의 및 기대 효과	74
3. 향후 연구	74

부록 1. 2021 개체명 분석 지침

부록 2. 개체 연결 말뭉치 구축 지침

참고문헌

<Abstract>

차 례

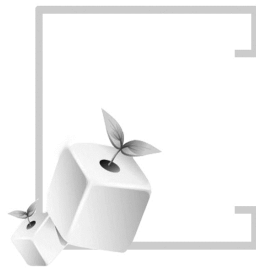
<표 차례>

<표 1> 사업 추진 경과	6
<표 2> 구축 도구별 설명	9
<표 3> ‘꽃’ 클러스터 문장 표	12
<표 4> 구축 작업자 교육 내용	24
<표 5> 말뭉치 구축 과정	29
<표 6> 개체명 부착 작업자 작업방식	30
<표 7> 평가 작업자 작업방식	30
<표 8> 개체명 연결 작업자 작업방식	31
<표 9> ‘21 신규 말뭉치 구축 및 수정 작업 내용	31
<표 10> ‘19 개체명 말뭉치 확장작업 내용	32
<표 11> 개체 연결 구축 작업 내용	33
<표 12> ‘21 평가 말뭉치 구축 및 수정 작업 내용	33
<표 13> 말뭉치 납품 결과	34
<표 14> 품질 측정 결과	36
<표 15> 개체명 사전 구조 예제	37
<표 16> 원시 말뭉치 JSON 형식	38
<표 17> 세부 분류 전체 빈도	40
<표 18> 세부 분류에 연결된 위키페이지 전체 빈도	43
<표 19> 가장 많이 연결된 위키피디아 페이지 (제목)	44
<표 20> 지침 회의 및 협의 내용	51
<표 21> 전문가 자문 내용 1	66
<표 22> 전문가 자문 내용 2	67
<표 23> 전문가 자문 내용 3	67
<표 24> 지침 회의 및 협의 내용 2	69

차 례

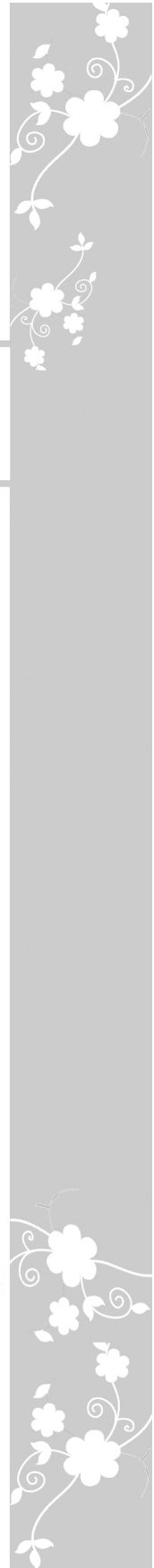
<그림 차례>

<그림 1> 사업의 목적	3
<그림 2> 사업 수행 기관	5
<그림 3> 케인 세부 기능 1	15
<그림 4> 케인 세부 기능 2	16
<그림 5> 와이즈돔 편의 기능	18
<그림 6> 케인 사용 방법 1	19
<그림 7> 케인 사용 방법 2	19
<그림 8> 케인 사용 방법 3	20
<그림 9> 와이즈돔 작업 방법 1	21
<그림 10> 와이즈돔 작업 방법 2	21
<그림 11> 와이즈돔 작업자 교육자료	22
<그림 12> 작업자 제출 서류	25
<그림 13> 작업자 공지사항 및 제출 현황	26
<그림 14> 사업 수행 조직	27
<그림 15> 상위 15개 카테고리 출현 빈도	38
<그림 16> 위키페이지가 연결된 빈도	42
<그림 17> 다의어 도식화	57
<그림 18> 동음이의어 도식화 1	57
<그림 19> 동음이의어 도식화 2	58
<그림 20> 상위어-하위어 도식화 1	58
<그림 21> 상위어-하위어 도식화 2	58
<그림 22> 전체어-부분어 도식화 1	59
<그림 23> 전체어-부분어 도식화 2	59
<그림 24> 전체어-부분어 도식화 3	59
<그림 25> 전체어-부분어 도식화 4	60
<그림 26> 동일지시적 표현	60



제 1 장

사업 개요



1. 사업 목적



<그림 1> 사업의 목적

‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업은 4차 산업혁명과 비대면 업무처리의 증가로 주목받고 있는 빅데이터 및 인공지능 기술 개발의 토대가 되는 언어 데이터의 필요성을 인지하여 활용도가 높은 개체명 분석과 개체 연결 분석 연구를 진행한다. 대용량·고품질의 개체명 및 개체 연결 말뭉치 데이터를 구축한 뒤 배포하여, 인공지능 언어처리 기술 개발 및 연구의 기반 자료로 활용될 수 있도록 한다. 이 과정을 통해 국내 표준화 및 참조, 평가 기반이 될 수 있는 자료들을 생성하고 배포함으로써 공공의 이익에 이바지하고 이후에 이어질 관련 연구의 방향성과 기준을 제시하고자 한다.

2. 사업 범위

본 사업의 수행 범위는 다음과 같다.

□ 개체명 구축

- 국립국어원에서 보유 중인 ‘문어/구어 원시 말뭉치 300만 어절’을 대상으로 개체명 세분류 150개에 속하는 개체명을 분석하고 표지를 부착한다.
- 국립국어원에서 보유 중인 ‘국립국어원 개체명 분석 말뭉치 (버전 1.0)’ 300만 어절을 대상으로 기존에 부착된 개체명 대분류 15개를 세분류 150개로

확장하여 개체명을 분석하고 표지를 부착한다.

- 분류 항목의 수가 많은 만큼 반복된 검증을 통해 오류율을 최소화하고, 데이터가 보유하고 있는 구축 프로세스 및 시스템을 적극 활용하여 분석한다.
- 정확률 97% 이상의 고품질 개체명 분석 말뭉치 구축을 목표로 한다.

☐ 개체명 구축 지침 재정비

- <2020년 개체명 분석 말뭉치 구축 지침 1.6버전> (2020년 개체명 말뭉치 연구 분석 보고서, 국립국어원, 2020)의 체계화 및 정밀화를 위한 재정비 작업을 진행한다.

☐ 개체 연결 정보 부착

- 개체명 분석 말뭉치 총 1,100만 어절을 대상으로 문서 내 PERSON, ARTIFACTS, ORGANIZATION, LOCATION, DATE 표지가 부착된 개체 표현이 의미상으로 지칭하는 개체를 지식베이스(위키피디아) 상에서 찾아 연결 정보를 부착한다.

☐ 개체 연결 지침 구축

- 개체 연결 작업을 위한 방법론 및 지침을 작성한다.
- 이전에 구축된 국립국어원의 개체명 말뭉치를 분석하여 개체명 대분류 15개 중 가장 높은 비중을 차지하거나 고유명사가 많이 포함되는 대분류 5개 (PERSON, ARTIFACTS, ORGANIZATION, LOCATION, DATE)를 선정하여 위키피디아 문서와 연결하는 방안을 기술한다.

☐ 개체명 사전 구축

- 개체명 및 개체 연결 정보가 부착된 말뭉치 1,100만 어절에서 개체 표현, 개체 유형, 개체 정보를 추출하여 개체명 사전을 구축한다.

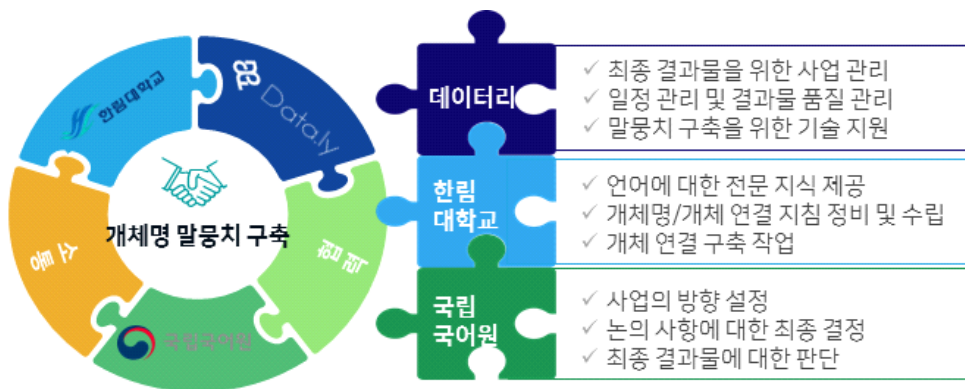
☐ 납품 자료의 품질 보증 및 보완 체계 수립

- 납품 자료 오류 발견으로 수정이 필요할 경우 하자 보수를 위한 체계를 설계한다.

- 개체명 및 개체 연결 말뭉치의 자동 검증, 전문가 검증, 통계적 검증을 통하여 말뭉치 분석 정확률 97% 이상을 달성한다.

본 사업팀은 사업 진행 과정 전반을 위 사업 범위에 맞추어 설정하였고 사업이 종료 기간 내에 성공적으로 마무리될 수 있도록 진행하였다.

3. 사업 수행 기관



<그림 2> 사업 수행 기관

본 사업의 참여 기관은 데이터리(Data.ly)와 한림대학교 두 기관이다. 대량의 고품질 개체명 및 개체 연결 말뭉치 구축 작업과 개체명 분석 지침 재정비 및 개체 연결 분석 지침 작성 작업을 효율적으로 수행하기 위해 각 기관의 전문성을 발휘할 수 있는 과업을 중심으로 역할을 분담하였다.

데이터리는 사업의 전반을 총괄하며 ‘기술팀’으로서의 역할을 수행하였다. 국립국어원으로부터 전달받은 원시 말뭉치를 작업자들에게 배포하기 전, 1차 구축을 위하여 자체적으로 보유한 개체명 자동 분석 시스템과 프로세스를 이용하여 개체명 표지를 부착한다. 1차 구축이 진행됨과 동시에 2차 구축을 위한 작업자들을 모집하여 개체명 교육을 진행한다. 1차 구축이 끝난 후, 작업 문서를 분배하기 전 일정한 양으로 나눈 다음, 개체명 도구 ‘케인(Kane)’에 적용이 가능한 형식으로 변환한다. 작업자들이 직접 문서를 확인하여 개체명 표지를 부착하는 2차 구축까지 마친 말뭉치는 자동화 시스템을 이용하여 오류 검출 및 수정 작업을 진행한다. 이런 과정을 거쳐 구축된 개체명 말뭉치 전체를 분석하여 개

체명 사전을 구축한다.

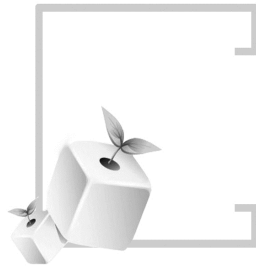
한림대학교는 국어학 전문가가 중심이 되어 ‘언어팀’의 역할을 수행하였다. 개체 연결 방법론 및 지침 수립을 위하여 기존의 개체 연결 연구를 분석하여 지침 구축을 위한 토대를 구성하였으며, 실용적인 지침 마련을 위하여 개체 연결 구축 작업을 병행하였다. 작업 시 발생했던 문제점이나 구체적인 설명이 필요한 부분을 파악하여 개체 연결 분석 지침을 작성하였다. 또한 작년에 작성된 <2020년 개체명 분석 말뭉치 구축 지침>의 재정비를 위해 국어학 전문가에게 자문해 지침에서 수정할 부분을 확인하고 국립국어원과 협의하여 최종적으로 도출된 결과를 지침에 반영하는 과정을 통해 개체명 분석 지침을 수정·보완하였다.

4. 사업 추진 경과

본 사업의 추진 경과는 다음과 같다.

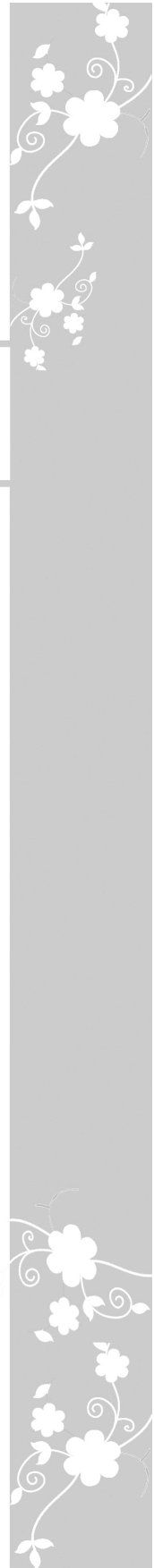
사업 추진 경과								
일련 번호	추진 내용	월						
		6	7	8	9	10	11	12
1	<ul style="list-style-type: none"> 말뭉치 구축 프로세스 및 구축 도구 준비 말뭉치 지침 분석 말뭉치 검증 시스템 준비 							
2	<ul style="list-style-type: none"> 말뭉치 구축 오류 탐지 오류 수정 개체명 지침 재정비 개체 연결 지침 작성 							
3	<ul style="list-style-type: none"> 전문가 평가 통계적 평가 분석 지침 평가 							
4	<ul style="list-style-type: none"> 말뭉치 구축 마무리 							
5	[산출물] <ul style="list-style-type: none"> 최종 산출물 							

<표 1> 사업 추진 경과



제 2 장

사업 수행 내용



1. 사업 수행 방안

‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업은 선행 사업보다 구축 말뭉치 작업량이 많이 증가하였고, 개체 연결 구축 및 지침 수립이라는 신규 과업이 추가되었다. 따라서 본 사업의 성공적인 수행 방안 마련을 위해 앞서 직접 수행하였던 개체명 사업 분석을 통하여 효율적이고 실제로 구현이 가능한 사업 수행 방안을 구성하였다. ‘구축 방안, 지침 방안, 평가 방안, 관리 방안’으로 나누어 각 방안에 대하여 기술하고자 한다.

1.1. 구축 방안

구축 작업을 위한 말뭉치 배포의 경우 늘어난 구축 말뭉치의 양과 그에 따른 구축 작업자의 증가로 인해 일정의 지연 없이 효율적으로 일을 처리할 방법이 필요했다. 이를 위해 클라우드 소싱 플랫폼 ‘와이즈돔(Wisedome)’을 사용하였다. 개체명 구축 도구 ‘케인(Kane)’과 병행하여 사용함으로써 한 작업 도구에 집중되어 발생하는 과부하 문제를 방지하고 작업을 분업화하여 체계적으로 배포하고 수합할 수 있도록 하였다.

구분	KANE	WISEDOME
작업 말뭉치	<ul style="list-style-type: none"> • '21 문어/구어 말뭉치 300만 어절 • '21 평가 말뭉치 • 개체 연결 말뭉치 1,100만 어절 • 기타 '19 '20 오류 사항 수정 	<ul style="list-style-type: none"> • '19 개체명 말뭉치 300만 어절
작업자	<ul style="list-style-type: none"> • '19 '20 개체명 작업에 참여했던 경험이 있는 자 • 교육 담당자로부터 대면 교육을 이수하고 지침의 전체 내용을 숙지한 자 	<ul style="list-style-type: none"> • 공개 모집을 통하여 선정된 자 • 교육 담당자로부터 비대면 교육을 이수하고 분담받은 개체명 분류에 대한 숙지가 완료된 자
작업 내용	<ul style="list-style-type: none"> • 원시 말뭉치에 150개 개체명 세분류를 기준으로 알맞은 표지 부착 • '21 문어/구어 말뭉치 평가를 위한 말뭉치 (전체의 10%)를 2년간 개체명 구축 작업을 한 전문가가 구축 • '19 '20 '21 개체명 말뭉치 1,100만 어절을 대상으로 5개의 대분류 (PER, AFW, ORG, LOC, DAT) 개체명 표지가 부착된 개체 표현에 해당 표현의 의미를 설명하는 위키피디아 페이지 주소 정보 부착 	<ul style="list-style-type: none"> • 기계를 통해 1차로 15개의 대분류에서 150개의 세분류 개체명으로 범위확장 • 15개 대분류를 3개씩 묶어 5개의 그룹에 분배하고, 작업자들은 해당 개체명 대분류 숙지 후 기계로 150개 개체명 표지를 부착한 개체명에 오류를 확인하는 작업 진행

<표 2> 구축 도구별 설명

1.2. 지침 방안

〈2020년 개체명 분석 말뭉치 구축 지침〉 업무를 담당하여 개체명 분석과 지침에 대한 이해도가 높은 전문가들이 올해 사업에도 참여해 지침을 재정비하였다. 선행 사업에서 수정되지 못했던 부분이나 일관성이 없는 부분 이외에 결과물을 바탕으로 추가되어야 할 사항 등을 분석하여 지침의 완성도를 높였다. 개체 연결 지침의 경우, 언어학 전문가들이 사업 시작 전부터 해당 연구를 진행하여 미리 초안을 작성하였다. 여기에 개체 연결 말뭉치 구축 작업을 병행하며 추가 기술 사항을 파악하여 이론적인 부분과 실용적인 부분을 모두 반영한 지침을 완성하였다.

1.3. 평가 방안

구축된 말뭉치의 성능 평가를 위해 전체 말뭉치 내 일관성 검사와 개체명 전문가들이 구축한 평가 말뭉치와의 비교를 통해 평가를 진행하였다. 일관성 검사의 경우 다수의 작업자에게 같은 문서를 배분하고 그 작업물을 비교하는 다면 평가 방법을 통해 확인하였으며 오류로 검출되는 부분은 재전달하여 수정하는 작업을 주 간격으로 반복하여 진행하였다. 또한 평가 기준을 만들기 위해 2년간 개체명 사업에 투입되어 구축 작업을 진행하며 역량을 인정받은 전문가들이 전체 말뭉치의 10%를 평가 말뭉치로 구축하였다. 국어원에서 직접 구축한 말뭉치 또한 평가의 기준으로 사용하였다. 평가 말뭉치와 다른 작업자들의 결과물을 비교하여 성능이 97%에 도달하는 것을 목표로 오류를 검출해내고 수정하는 작업을 반복하여 진행하였다.

1.4. 관리 방안

7개월이라는 사업 기간 내에 많은 과업을 성공적으로 수행하기 위해 선행 사업의 경험을 바탕으로 구체적으로 사업 일정을 계획하였다. 개체명 구축의 경우 주 단위로 수립한 계획 일정에 따라 차질 없이 진행하였으며, 개체 연결 구축은

월 단위로 계획하여 배포 및 수합 작업을 원활히 관리하였다. 또한 주간 보고와 월간 보고를 통해 사업의 척도를 국립국어원에 보고하였다.

2. 사업 수행 환경

2.1. 작업 공정

2.1.1. 사업 추진 전략 및 지원 체계

본 사업은 말뭉치 및 사전 구축으로 다음의 목표를 가지고 있다. 첫 번째는 15개 범주로 분석된 2019년 개체명 말뭉치를 150개의 개체명 범주로 확장하는 것, 두 번째는 2020년 개체명 말뭉치를 변경된 지침에 맞게 수정하는 것, 세 번째는 2021년 개체명 말뭉치를 구축하는 것이다. 네 번째는 19년, 20년, 21년 개체명 말뭉치에 개체 연결 정보를 부여하는 것이고 마지막 다섯 번째는 개체명 사전을 구축하는 것이다.

2019년 말뭉치 300만 어절(약 37만 문장), 2020년 말뭉치 500만 어절(약 134만 문장), 2021년 말뭉치 300만 어절(약 41만 문장)의 대용량 말뭉치를 구축하는 것으로 구성된 팀 간의 소통과 사업 수행 방향을 설정하는 것이 중요하다. 이를 위하여 개체명 말뭉치는 크게 지침 분석팀(검증팀)과 말뭉치 구축팀(구축팀)으로 구성되고 말뭉치 구축팀은 신규 말뭉치 구축팀, 말뭉치 검증팀, 말뭉치 보완팀, 말뭉치 확장팀으로 세분화한다. 신규 말뭉치 구축팀은 2021년 말뭉치를 담당하고 검증팀은 2021년 말뭉치 중 10%에 해당하는 문장을 구축하여 평가 항목 등을 담당하며 확장팀은 2019년 15개의 상위 범주를 150개로 확장하는 작업을 담당한다.

개체 연결 또한 검증팀과 구축팀(신규 말뭉치 구축팀)을 구성하였고, 개체명과 동일하게 운영한다. 개체 연결은 지침에 대한 정의를 사업 초반에 진행하고 구축팀이 말뭉치 구축을 진행하며 지침을 보완한다. 개체명 사전은 개체명과 개체 연결 말뭉치를 바탕으로 개체명 정의, 유사어 등의 결과물을 통합하여 구성한다.

이처럼 본 사업 목표를 달성하기 위하여 구축팀과 검증팀이 유기적으로 작업할 수 있도록 지원이 필요하다. 구축팀은 지침에 맞게 말뭉치를 구축하는 것이 목표이고 검증팀은 지침의 보완, 수정 및 구축팀의 구축 결과물에서 수정이 필요한 항목 또는 분석 누락 항목 등을 전달하여 말뭉치의 품질을 향상시키는 역할을 한다. 검증팀과 구축팀의 협업을 지원하기 위해 말뭉치 구축 도구를 제공하고 먼저 학습된 전문가의 지식을 통해 구축팀의 작업을 보조하여 검증팀의 피드백을 반영하는 역할을 한다. 전문가 지식은 기존의 2019년과 2020년에 구축한 개체명 데이터를 이용한 사전 형식의 지식이다. 예를 들어 ‘청와대’가 어떤 개체명 범주로 작업되었는지 알 수 있도록 2019년, 2020년 검증팀의 구축 말뭉치를 활용하여 개체명의 빈도와 문맥 정보를 저장한 것이다. 검증팀 피드백은 구축팀의 작업 결과와 전문가 지식의 개체명 빈도 및 문맥 정보를 비교한 것으로 일치도가 낮으면 구축팀에게 재확인을 요청하였다.

2.1.2. 전처리 도구

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

두 번째는 유사한 개체명이 있는 문장들을 군집화(Clustering, 클러스터링)하는 기능이다. 군집화를 통하여 유사한 개체명을 포함한 문장들이 하나의 클러스터로 구성되는데 예를 들면 아래<표 3>과 같이 ‘꽃’과 관련된 문장들이 하나의 클러스터로 구성된다.

```

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

```

<표 3> ‘꽃’ 클러스터 문장 표

한 클러스터 내에 있는 유사한 개체명을 가진 문장들을 다수의 작업자가 주

석 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.

2.1.3. 점진적 처리 프로세스

개체명 말뭉치 작업은 전처리 도구를 통해 얻은 클러스터 결과물을 작업자와 시스템에 분배하여 주석 작업을 진행한다. 클러스터를 기반으로 분배를 하는 이유는 같은 클러스터 내에 표현된 개체명은 같은 범주를 가질 확률이 높기 때문이다. <표 3>과 같이 ‘꽃’과 관련된 클러스터 내에서 ‘핑크 튜립’, ‘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, 동일하지 않은 문장에 대해 주석 작업을 진행한다. 이후 주석 작업이 완료된 문장은 클러스터를 기준으로 분류하여 오류를 탐지하고 탐지된 오류를 기반으로 재수정하거나 교육을 시행한다.

이러한 과정을 여러 번 반복하는 것을 점진적 처리 프로세스로 정의하고 이를 통해 말뭉치 구축을 위한 지침을 정리하고 대량의 개체명 분석 말뭉치의 품질을 높인다.

개체 연결 말뭉치의 점진적 처리 프로세스는 클러스터 결과물 외 개체 연결 대상(PERSON, ARTIFACTS, ORGANIZATION, LOCATION, DATE)이 되는 개체명의 통계적 정보를 통해 개체 연결을 먼저 작업할 개체명과 문장을 선택한다. 2019년, 2020년 개체명 말뭉치는 개체명이 부착되어 있으므로 개체 연결 대상이 되는 개체명이 포함된 문장과 클러스터 결과물과 참고하여 먼저 주석 작업을 수행할 문장을 선택하는 것이다.

선택된 문장을 기반으로 개체 연결 주석 작업을 수행하면서 지침을 보완하고,

주석 작업이 완료되면 개체 연결 주석 작업이 되지 않은 개체명의 빈도 및 클러스터를 참조하여 추가 문장들을 선택하는 방법으로 진행한다. 이 방법은 개체명의 점진적 프로세스와 동일하지만 상위 15개의 범주가 아닌 대상이 되는 5개의 범주로 좁혀 처리하는 방식이다.

2.1.4. 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 실수가 잦아져 오류 발생 가능성도 커진다.

이 문제를 해결하기 위해 변경된 지침의 반영이 필요한 부분을 검색하고, 일치되는 부분을 말뭉치 구축팀 책임자가 검토하고 최종적으로 확정된 부분을 일괄적으로 반영함으로써 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.

2.2. 구축 도구

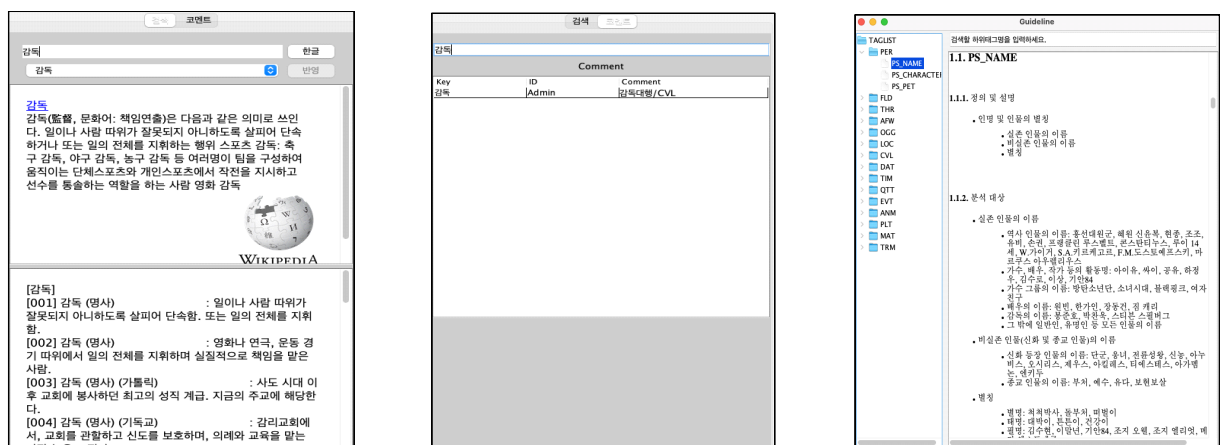
말뭉치 구축 및 연결 작업 시 시간 단축과 정확성을 높이기 위하여 개체명 부착 및 수정 프로그램인 ‘체인’과 ‘와이즈돔’을 사용한다. 두 구축 도구는 데이터리가 자체 제작하였다.

먼저 ‘체인’은 개체명 말뭉치 구축과 개체 연결 말뭉치 구축에 모두 사용할 수 있다. 개체명의 주요 기능은 작업자가 문장에서 개체명 구간을 선택하고 개체명 분류 체계에 따른 150개의 세부 분류 중 적절한 범주를 선택하는 것이다. 그리고 개체 연결에서는 개체명을 선택할 시 개체명을 위키피디아 제목을 기준으로 검색하고, 유사도가 높은 위키피디아 후보를 제시한다. 제시된 후보 또는 추가 검색을 통하여 올바른 위키피디아 페이지를 선택하는 방법으로 말뭉치를 구축한다. 문서는 데이터리가 관리하는 서버를 통해서 각 작업자에게 배분하고, 작업자들은 구축 도구를 통해 문서를 전달받아 작업 후 서버로 업로드한다.

말뭉치 구축 작업은 한 문서를 한 명이 작업하는 것이 아니라 여러 개의 문서를 여러 명이 작업하므로 일관성 있는 작업 결과물이 중요하다. 이를 위하여 ‘케인’ 도구에서는 다음과 같은 기능을 제공한다. 첫 번째는 코멘트 기능이다. 작업 창에서 개체명을 선택하거나 코멘트 검색창에 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 의미를 확인하는 일이 많다. 작업자가 선택하거나 입력한 개체명의 사전 정보를 자동으로 ‘위키피디아’와 <우리말샘>에서 검색하여 제시함으로써 작업자에게 편의를 제공한다. 개체 연결에서는 위키피디아 내용을 중심으로 볼 수 있고, 보이는 페이지를 개체명에 연결할 수 있다.

마지막 기능은 가이드라인 기능이다. 개체명 범주가 다양하고 지침의 내용에 따라 결정되는 항목이 많으므로 참고해야 하는 경우가 많이 발생한다. 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.



<그림 3> 케인 세부 기능 1

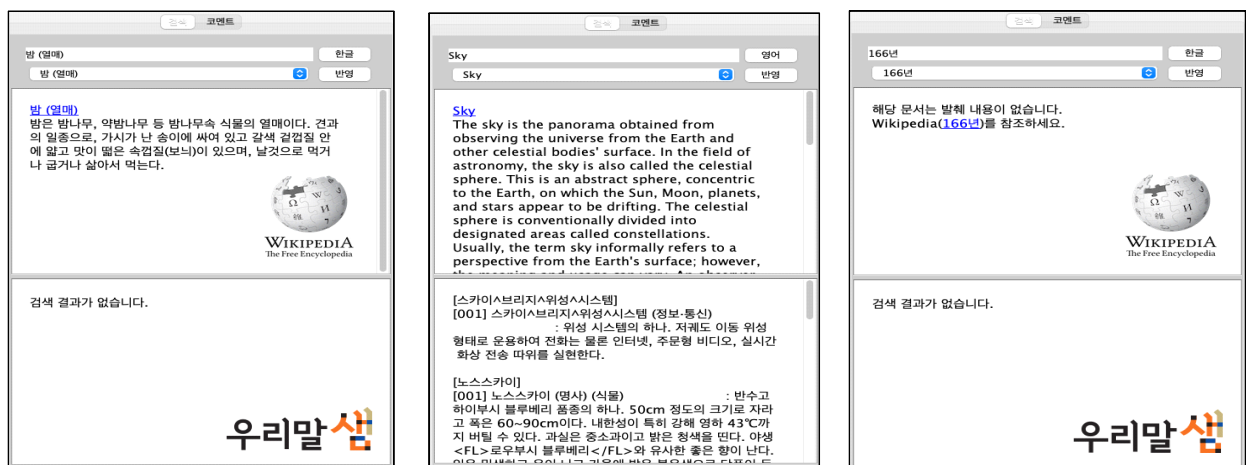
개체 연결 작업은 선택된 개체명에 대해 유사도가 높은 위키피디아 페이지(문서) 후보를 제시한다. 제시된 위키피디아 후보 중 올바른 항목이 있으면 연결하

고 또는 제시된 후보에 연결하고자 하는 항목이 없다면 작업자가 직접 검색을 통해 위키피디아 페이지를 찾아 연결하는 기능을 제공한다. 개체 연결을 위하여 제공하는 기능은 다음과 같다.

첫 번째는 검색 기능이다. 개체명의 동음이의성과 다의성으로 인해 유사한 제목이지만 다른 내용이 다양하게 존재한다. 검색 기능으로 추천 목록을 제공하여 목록에서 적합한 대표 페이지를 빠르게 찾을 수 있는 기능을 제공한다.

두 번째는 바로 가기 기능이다. 개체명을 검색한 검색 결과 창에 보이는 내용만으로는 개체 연결을 결정하기에 부족할 수 있다. 결과 창에 검색 결과와 같이 해당 원문을 확인할 수 있는 위키피디아 정보를 제공하기 때문에 작업자는 내용을 빠르게 확인할 수 있다.

세 번째는 위키피디아 언어별 검색 기능이다. 위키피디아는 언어별로 따로 관리되고 있다. 또한 본 사업에서는 한국어 위키피디아를 먼저 연결하는 것이 목표이므로 한국어 위키피디아에서 먼저 유사도를 검색하여 후보를 제시한다. 하지만 한국어 위키피디아 정보는 영어 위키피디아보다 정보가 적으므로 한국어에 없는 경우 영어 위키피디아를 검색할 수 있는 기능을 제공한다. 이러한 경우 사전 원어 정보와 외래어 표기 용례를 이용하여 영어 문서를 연결할 수 있도록 영어 검색 기능을 제공한다. 마지막 기능은 문서 연결 제한이다. 대부분의 개체 연결에 연결된 위키피디아 페이지는 해당 개체명의 정보를 충분히 포함하고 있는, 즉, 하나의 개체에 대한 설명 페이지를 사용한다. ‘동음이의어 문서’ 또는 ‘넘겨주기 문서’와 같이 적합하지 않은 페이지는 연결을 제한하여 작업자가 연결할 위키피디아 페이지를 결정하는 데에 참고할 수 있는 기능을 제공한다.



<그림 4> 케인 세부 기능 2

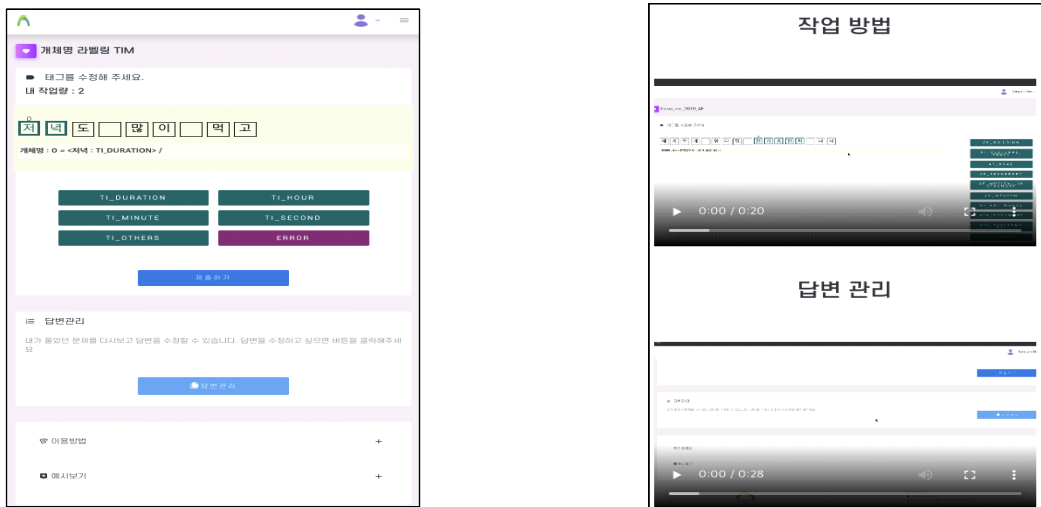
그 외 제공하는 기능 및 장점은 ‘케인’을 사용함으로써 작업자들은 화면 속 작업 창에서 마우스 클릭으로 개체명 작업을 수행하므로 입력 오류를 줄일 수 있고, 위에서 설명한 다양한 편의 기능이 포함되어 있으므로 부착 시간과 오류를 줄이는 데 효과적임을 사업을 통해 확인하였다. 그리고 작업 창에서 개체명을 선택하는 경우 오른쪽 창에 해당 개체명에 대한 코멘트와 사전 정보를, 별도 창으로 지침 정보를 실시간으로 나타내어 개체명 부착 작업의 효율을 높인다. 개인적으로 작업이 이뤄지지만, 공동 작업자들이 다른 작업자들의 결과물을 확인할 수 있어 협업이 가능한 것이다.

구축 도구 ‘와이즈돔’은 웹 기반의 클라우드 소싱 플랫폼이다. ‘케인’과 달리 ‘와이즈돔’은 저작 환경을 웹을 통해 다양하게 표현 가능하며 다수 작업자의 결과물을 확인할 수 있다. 본 사업에서는 2019년에 구축된 국립국어원 개체명 분석 말뭉치(버전 1.0) 300만 어절을 대상으로 하는 태깅 확장 작업에 활용하여 상위 15개 개체명 범주로 주석 작업이 완료된 말뭉치를 150개의 세부 범주로 확장하였다.

플랫폼에 15개 범주의 프로젝트를 생성하였고 각 프로젝트는 해당 범주의 개체명이 포함된 문장을 작업할 수 있다. 예를 들어 ‘PERSON’ 프로젝트에 포함된 문장은 상위 범주로 표시된 인명을 세부(PS_NAME, PS_CHARACTER, PS_PET) 범주로 분류하는 작업을 수행하는 것이다. 각 개체명에 대해서 한 명의 결과를 얻는 것이 아니라 다수 작업자 결과물을 얻어 일치율을 확인 후 해당 개체명의 세부 개체명 범주를 확정하므로 일관성 및 정확도를 높일 수 있다. 세부적으로 ‘와이즈돔’에서는 다음과 같은 기능을 제공한다. 첫 번째는 ‘제약이 없는 작업환경’이다. ‘와이즈돔’은 웹 기반 플랫폼으로 온라인상에서 공간과 시간의 제약 없이 작업자가 어디서든 원하는 시간과 방법으로 작업을 할 수 있는 작업자 친화적인 환경을 제공한다. 두 번째는 ‘집중력을 높이는 작업방식’이다. 한 화면에 여러 문장을 보면서 작업하는 방식이 아니라 한 화면에 한 문장만을 보며 작업하는 방식이다. 작업자가 작업하는 문장마다 집중력을 발휘할 수 있는 환경을 만들어 오류를 줄일 수 있도록 했다. 세 번째는 ‘전체적인 문맥을 확인하는 기능’이다. 문맥에 따라 개체명의 의미가 달라진다. 이에 한 문서 내에 같은 개체명이 2개 이상의 개체명 표지를 가지는 경우를 추출하여 문서 단위로 수정할 수 있도록 하였다.

마지막은 ‘작업 방법 동영상’이다. 작업자마다 작업 방법이 다른 것은 작업 결과물의 일관성에 영향을 줄 수 있으며 효율적이지 못하다. 프로젝트마다 작업 방법을 동영상으로 제공함으로써 작업자들이 작업 방법에 대해 빠르게 이해하고 일관성 있는 작업을 가능하게 하여 더욱 좋은 결과물을 기대할 수 있다.

‘와이즈돔’을 사용함으로써 작업자들은 웹브라우저 화면에서 선택만으로 개체명 작업을 수행하므로 입력 오류를 줄일 수 있다. 또한 위에서 설명한 다양한 편의성을 제공하므로 부착 시간과 오류를 줄이는 데 효과적이다. 또 작업 화면, 답변 관리 화면과 작업 방법 영상을 한 화면에서 확인할 수 있도록 해 작업자들은 빠르고 간편한 작업이 가능하다.



<그림 5> 와이즈돔 편의 기능

본 사업에서 ‘케인’과 ‘와이즈돔’의 사용은 개체명 부착에 대한 오류 감소 및 시간을 단축하는 데 아주 중요한 역할을 했다. 본 도구들을 이용함으로써 구축 작업의 정확도와 능률이 월등히 높아졌으며 사업 진행에 큰 도움이 되었다.

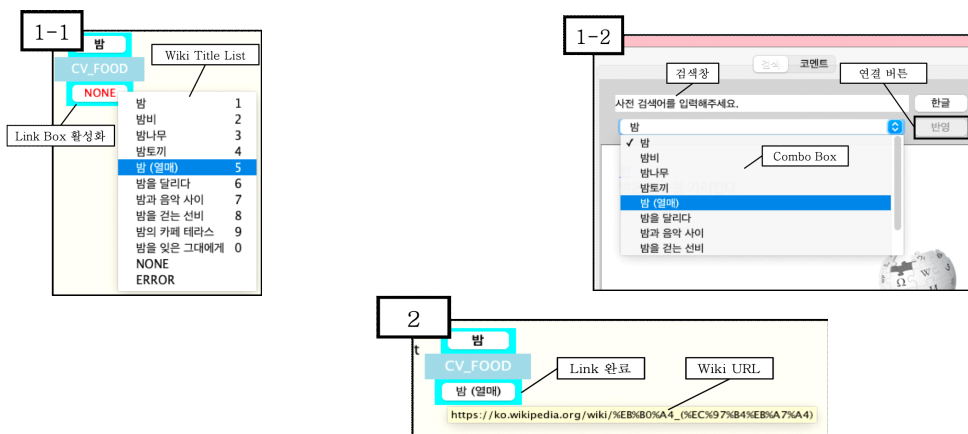
2.2.1. 말뚝치 구축 및 연결 작업 방법

□ 케인(Kane)



<그림 6> 케인 사용 방법 1

태깅 방법은 <그림 6>의 순서와 같다. ‘케인’에서 작업할 문서를 연 다음, 문서에 포함된 개체명을 선택하여 개체명에 부합하는 태그를 선택한다. 개체명을 선택하는 방법은 문장에서 개체명에 해당하는 구간을 드래그하거나, 대상 어절 전체가 개체명일 때는 더블 클릭을 하면 된다. 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 세부 분류 중 하나를 선택할 수 있다. 태그가 부착된 개체명은 문장 내에서 개체명 작업 여부를 직관적으로 확인할 수 있도록 개체명이 아닌 것과 구분하여 표시된다.



<그림 7> 케인 사용 방법 2

개체 연결 방법은 <그림 7>의 순서와 같다. ‘케인’ 상단 메뉴바에 보기에서 ‘연결 박스 활성화’를 할 수 있다. 활성화가 되면 태그가 부착된 개체명의 태

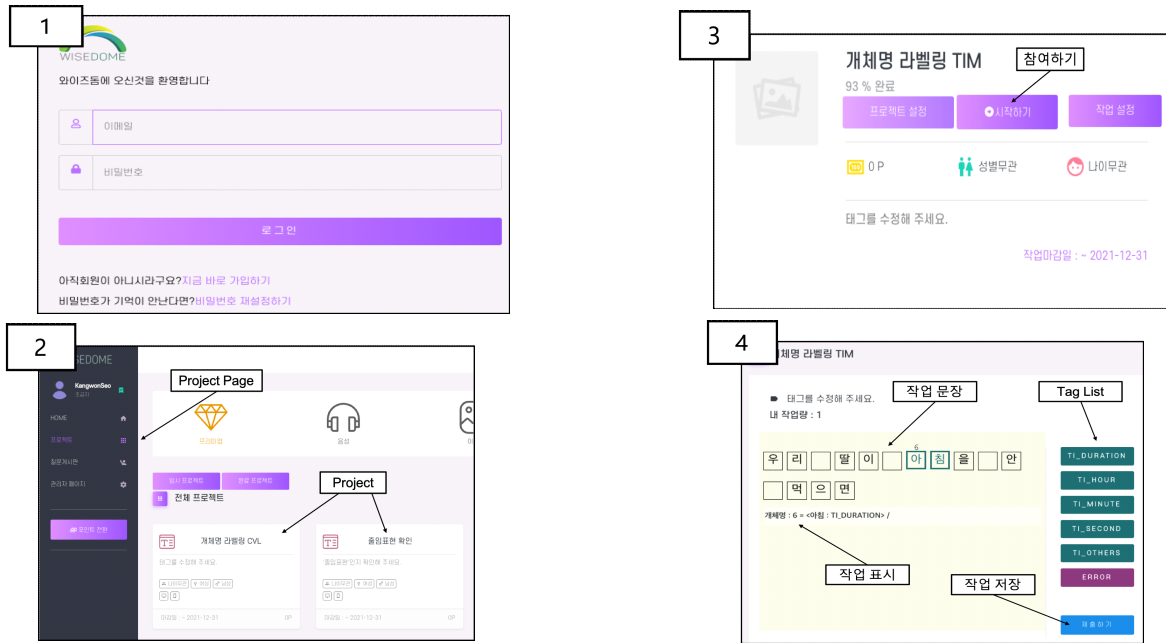
그 박스 아래에 연결 박스가 표시된다. 표시된 연결 박스에 개체명에 부합하는 위키피디아 문서 제목을 기준으로 한 대표 페이지 URL 주소를 선택하는 것이다. 위키피디아 문서 선택 방법은 연결 박스를 이용하는 방법과 검색창을 이용하는 방법이 있다. 연결 박스를 이용하는 방법은 연결하고자 하는 개체명의 연결 상자를 클릭하면 ‘추천 문서 목록’을 보여주는데, 개체명에 해당하는 위키피디아 문서를 찾아 선택하면 된다. 검색창을 이용하는 방법은 ‘케인’ 화면 오른쪽에 있는 검색창에 개체명을 검색하면 검색창 아래에 있는 콤보 박스(Combo Box)에 보이는 ‘추천 문서 목록’에서 적합한 위키피디아 문서를 찾아 선택한 후 ‘연결 버튼’을 클릭하면 된다. 선택한 문서가 ‘없는 문서’, ‘발췌 내용이 없는 문서’, ‘동음이의어 문서’ 일 경우 연결할 수 없다. 연결 결과는 선택한 개체명의 연결 박스에 보이는 문서 제목과 툴 팁으로 보이는 URL 주소로 확인할 수 있다.



<그림 8> 케인 사용 방법 3

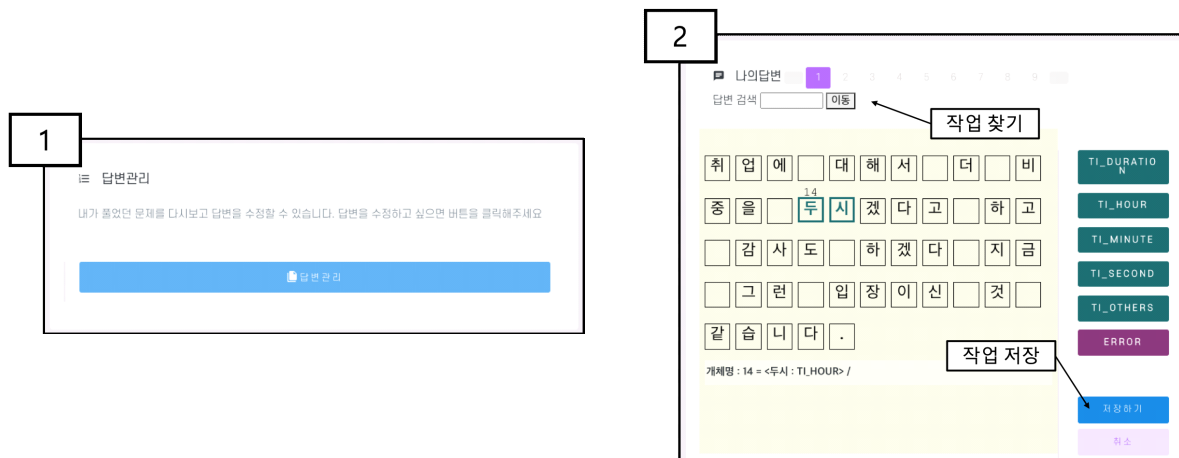
개체명 구간이 잘못 선택된 경우 더블 클릭을 하여 제거할 수 있다. 또한 현재 작업 문서 내에 동일 개체명이 여러 개 나타날 경우 일괄적으로 태깅, 검색, 변경, 삭제할 수 있는 기능을 제공하여 작업자가 빠르고 간단하게 작업할 수 있다.

□ 와이즈돔(Wisedom)



<그림 9> 와이즈돔 작업 방법 1

작업 방법은 <그림 9>의 순서와 같다. 작업자는 ‘와이즈돔’에 접속 후 로그인하고 프로젝트 페이지에서 담당 프로젝트를 찾아 들어가 ‘시작하기’를 누르면 참여할 수 있다. 프로젝트가 시작되면 화면에 보이는 문장에 포함된 개체명을 선택하여 개체명에 부합하는 태그를 선택하는 것이다. 개체명을 선택하는 방법은 문장에서 개체명에 해당하는 구간의 처음절과 끝음절을 찾아 선택하면 된다. 개체명 선택 후 세부 분류 버튼 중 하나를 선택할 수 있다. 작업 내용은 문장 아래에 표시되며, 태그가 부착된 개체명은 문장 내에서 개체명 작업 여부를 직관적으로 확인할 수 있도록 개체명이 아닌 것과 색깔로 구분하여 표시된다. 완료되면 제출하기를 눌러 답변을 저장한다.



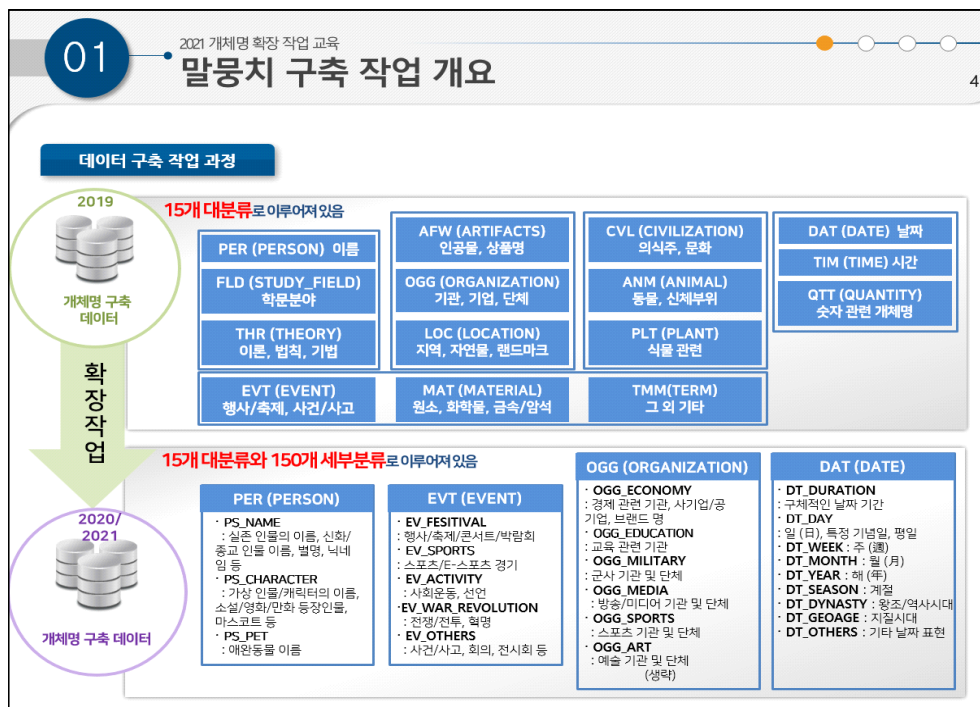
<그림 10> 와이즈돔 작업 방법 2

제출한 답변이 잘못되었으면 작업화면 아래 ‘답변 관리’에서 수정할 수 있다. 답변 관리 버튼을 눌러 답변 관리화면을 열고 수정할 문장을 찾아 재작업 후 저장하기를 누르면 된다. 수정 방법은 작업 방법과 같으며 답변 검색 기능을 제공하여 작업자가 빠르고 간단하게 작업 내용을 확인할 수 있다.

2.3. 작업자 교육

본 사업의 경우 짧은 사업 기간 내에 대용량 말뭉치를 구축하여야 하는 만큼 많은 작업자가 투입되었다. 다수의 작업자가 개체명 작업에 대해 충분히 숙지할 수 있도록 개체명 도구별로 교육 집단을 나누어 각 집단의 특성에 맞추어 맞춤 교육을 준비하였다.

우선 ‘와이즈돔’ 작업자의 경우 개체명에 대한 이해와 그룹별로 분담받은 3개의 대분류에 대한 숙지가 필요했으므로 총 5개의 그룹으로 나누어 그룹별로 비대면 교육을 진행하였다. 추가 교육이 필요한 작업자가 있거나 지침이 수정되면 상시로 비대면 교육을 진행하였다.



<그림 11> 와이즈돔 작업자 교육자료

‘케인’ 작업자는 원시 말뭉치에 150개 세분류 개체명 표지를 부착하는 고난도의 작업을 진행하는 만큼 개체명 지침에 대한 깊은 이해가 필요하다. 따라서 ‘케인’ 작업자들은 모두 대면으로 사업 및 개체명에 대한 전반적인 안내와 150개 세분류에 대한 교육을 진행하였다. 이후 개인별 작업물을 확인하고 추가 교육이 필요한 작업자의 경우 비대면 화상 회의를 통해 1:1 교육을 진행하였으며 전체 작업자를 대상으로 좀 더 숙지가 필요한 부분을 문서로 정리하여 전달하는 등 수시로 피드백을 제공하고 재교육을 진행하였다.

진행된 교육 일정과 내용은 다음과 같다.

일시	차수	진행방식	내용
21.06.28.	1차	대면 교육 (Kane)	<ul style="list-style-type: none"> • ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업에 대한 전반적인 안내 • 개체명 구축 도구 ‘Kane’ 안내 • 개체명 지침 교육 (세분류 150개 분류에 대한 설명) • 세분류 150개 분류를 이용한 태깅 작업 실습 • 보안서약서, 참여확약서, 근로 계약서 작성
21.06.29.	2차	대면 교육 (Kane)	<ul style="list-style-type: none"> • ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업에 대한 전반적인 안내 • 개체명 구축 도구 ‘Kane’ 안내 • 개체명 지침 교육 (세분류 150개 분류에 대한 설명) • 세분류 150개 분류를 이용한 태깅 작업 실습 • 보안서약서, 참여확약서, 근로 계약서 작성
21.06.29.	3차	대면 교육 (Kane)	<ul style="list-style-type: none"> • 평가 말뭉치 구축 작업자 대상 교육 진행 • 2020년 개체명 분석 지침 주요 변경사항 안내 • 2개년도 질의응답 정리 자료를 통한 오류 검출 내용 안내 • 보안서약서, 참여확약서, 근로 계약서 작성
21.06.29 ~ 21.07.12	4차	실습 교육	<ul style="list-style-type: none"> • 2021년 개체명 구축 작업자 대상 교육문서 실습 • 작업자가 제출한 문서를 정답과 비교하여 오류 산출 후 재전달하여 작업자가 확인하고 수정하는 방식
21.07.13 ~ 21.07.19	5차	실습 교육	<ul style="list-style-type: none"> • 2021년 개체명 구축 작업자 대상 교육문서 실습 • 작업자가 제출한 문서를 정답과 비교하여 오류 산출 후 재전달하여 작업자가 확인하고 수정하는 방식
21.08.17.	6차	ZOOM을 이용한 비대면 교육 (Wisedome)	<ul style="list-style-type: none"> • ‘Wisedome’ 이용 방법 및 개체명 확장작업 방법 안내 • 그룹 A(DAT, TIM, QTT), B(CVL, ANM, PLT), D(PER, FLD, THR) 별 담당 개체명 분류 안내 • 질의응답 및 주의 사항 전달
21.08.18.	7차	ZOOM을 이용한 비대면 교육 (Wisedome)	<ul style="list-style-type: none"> • ‘Wisedome’ 이용 방법 및 개체명 확장작업 방법 안내 • 그룹 C(AFW, ORG, LOC), E(EVT, MAT, TRM) 별 담당 개체명 분류 안내 • 질의응답 및 주의 사항 전달
21.08.27.	8차	ZOOM을 이용한 비대면 교육	<ul style="list-style-type: none"> • 국어원 시스템 작업 이후 제출물 성능이 저조한 작업자를 대상으로 개인 비대면 교육 진행 • 국어원 시스템 작업 결과와 개인 작업 결과 비교하여 오류 수정
-	-	웹 공지	<ul style="list-style-type: none"> • 이 외 지침의 변경사항이 있을 시 교육문서 작성 후 수시로 작업자들에게 공지

<표 4> 구축 작업자 교육 내용

1~3차 교육에서는 ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업에 대한 소개와 구축 작업에서 쓰일 ‘케인’ 도구에 대한 안내가 이루어졌다. ‘케인’을 이용하기 위한 아이디 생성 및 개인별 작업자 번호 안내, 작업물

업로드-다운로드 방법, 표지 부착 방법, 편의 도구 이용 방법 등 세부적인 교육을 진행하였다. 지침 교육은 먼저 개체명의 태깅 범주와 대상에 대하여 교육하였으며, 세분류 150개의 분류에 대하여 진행하였다. 교육 담당자의 구축 작업 시연을 통해 작업에 대한 이해도와 개체명 분류를 익히고 작업에 참여하는 모든 작업자들은 근로 계약서, 보안서약서를 제출하였다.

The figure shows two forms side-by-side. The left form is titled '보안서약서' (Security Agreement) and the right form is titled '단기근로계약서' (Part-time Worker Agreement). Both forms have a header section with personal information (성명, 성, 성명, 생년월일, 주민등록번호, 주소, 연락처, 이메일, 직책, 직책명). Below the header, there are several sections of text, including a '개인정보 수집 및 이용 동의' (Personal Information Collection and Use Consent) section. The forms are dated 2021년 6월 28일 and have a signature line for '성명: [Blank]'. The right form also has a section for '사업장' (Workplace) and '근로자' (Worker) information.

<그림 12> 작업자 제출 서류

4~5차 교육에서는 실습 교육을 진행하였다. 세분류 150개에 대한 숙지가 단시간 내에 불가할 것으로 판단하고 첫 3주까지는 교육문서 작업을 통해 작업자들의 숙련도를 증진시켰다. 교육문서는 작년 제출하였던 개체명 말뭉치 중 검수가 완료되어 정답이 존재하는 문서이다. 작업자들이 작업한 문서를 정답과 비교하여 오류 리포트를 출력한 다음, 이를 작업자들이 전달받아 작업 문서를 수정하는 방식으로 교육을 진행하였다.

6~7차 교육은 비대면으로 진행되었으며 ‘와이즈돔’ 이용 방법과 개체명 확장작업에 대한 전반적인 설명, 개체명의 태깅 범주와 대상에 대하여 교육하였고 그룹별 담당 개체명 분류에 대한 설명을 진행하였다.

8차 교육은 국어원 검증 시스템 작업 이후 작업자별 성능 결과를 확인하여 성적이 저조한 작업자를 대상으로 1:1 비대면 재교육을 시행하였다. 작업자가 구축한 작업 결과와 교육 담당자가 구축한 작업 결과를 비교하며 오류 수정을 진행하였으며 작업자가 주로 실수하는 부분을 분석하여 재교육하였다.

2.4. 사업 수행 조직



<그림 14> 사업 수행 조직도

주 수행 기관은 데이터리(Data.ly)이며, 공동 수행 기관으로 한림대학교가 참여했다. 각 기관의 장점을 살려 데이터리는 ‘기술팀’으로서 전체 사업의 길잡이 역할과 기술 및 기능적인 부분을 맡았다. 한림대학교는 ‘언어팀’으로서 지침 수립 및 개체 연결 말뭉치 구축 등 언어적인 부분을 맡았다.

‘기술팀’의 하위 부서로는 사업 지원팀, 기술 개발팀, 기술지원팀, 말뭉치구축팀이 있다. 사업 지원팀은 본 사업의 총괄을 맡는다. 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 공동사의 의견을 종합하여 기술팀이 대표로 국립국어원과 소통한다. 또한 사업의 진행 방향과 일정에 차질이 없도록 사업 주관 기관인 국립국어원에 사업의 진행 척도를 주기적으로 보고하는 역할을 한다.

기술 개발팀은 구축 작업에 사용될 도구 ‘케인’과 클라우드 소싱 플랫폼 ‘와이즈돔’의 개발 및 관리 업무를 맡으며, 작업물의 배포 및 수합과 오류 검출 등 구축 프로세스를 전담한다. 말뭉치의 성능 확인과 버전별 관리를 통해 결과물 제출에 차질이 없도록 한다. 사업 진행 내내 도구의 성능 확인 및 웹 환경 관리를 맡으며 외부의 접근 및 보안에 대한 관리를 맡아 사업이 철저한 보안 속

에서 이루어지도록 힘쓴다.

말뭉치 구축팀은 실질적으로 말뭉치 구축 작업을 진행하는 부서로서 구축 작업자들과 그들을 관리하는 담당자로 구성되어 있다. 구축 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. 담당자는 작업자들의 구축 결과물 검증과 작업자 교육 및 관리의 업무를 맡으며 사전에 구축 도구를 테스트하여 문제점과 보완점을 기술개발팀에 보고하여 구축 도구 기능 향상에 이바지한다.

마지막으로 기술지원팀은 과거 선행 사업 내용을 분석하여 문제점을 검토한다. 도출된 사항은 본 사업의 수행 방안을 수립할 때 반영한다. 즉각적으로 사업 진행 내용 등을 정리하며 개발된 도구와 프로세스 등을 미리 테스트해본 후 피드백을 전달한다. 또한 사용되는 서버와 웹 등의 환경을 관리하여 사업이 원활히 진행될 수 있도록 보조한다.

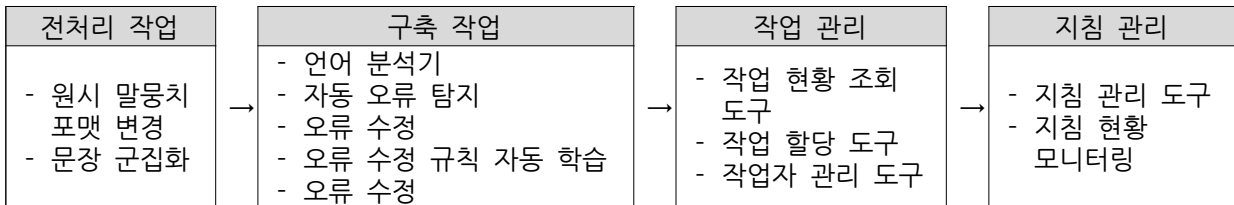
‘언어팀’은 지침 분석팀과 개체 연결팀으로 이루어져 있다. 지침 분석팀은 작년 개체명 사업을 진행하며 <2020년 개체명 분석 말뭉치 구축 지침> 수정 작업에 참여하였던 전문가들로 구성되었다. 이들은 개체명 지침을 분석하여 세분류별 정의 및 예제의 적합성과 일관성을 확인하고, 작년 사업에서 완벽히 정립하지 못했던 부분을 찾아 수정하고 보완하는 지침 재정비 작업을 진행한다. 지침 재정비 과정에서는 국어학 전문가들의 의견을 반영하였으며, 자문 결과와 내부 검수 결과를 토대로 국립국어원과 협의하여 수정 작업을 진행하였다.

개체 연결팀은 이번 사업에 신규 추가된 개체 연결 과업을 전담한다. 언어학 전문가들은 사전 연구를 통해 개체 연결 구축을 위한 지침의 초안을 작성하여 개체 연결 구축 작업의 진행 방향을 설정한다. 국립국어원과의 협의를 거치며 지침이 수립된 후, 전문가들은 구축 작업을 직접 진행한다. 구축 작업 시 설명이 부족하여 기준이 모호한 부분이나 추가할만한 예제들은 내부 회의를 통해 지침에 반영되어 지침의 명확성과 구체성을 높인다.

‘기술팀’과 ‘언어팀’ 그리고 그의 하위 부서들은 각자 맡은 역할에 최선을 다하여 사업을 수행하였고 유기적으로 연결된 부서들의 역량이 모여 사업이 성공적으로 마무리될 수 있었다.

3. 말뭉치 구축 및 납품

3.1. 말뭉치 구축



<표 5> 말뭉치 구축 과정

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

첫째, 전처리 작업에서는 원시 말뭉치의 포맷을 변환하여 문장 단위로 자동 분리하는 도구와 유사한 개체명이 있는 문장을 모아(군집화) 작업 대상을 선정하는 도구를 이용하여 작업자들에게 전달하기 위한 문서를 정제한다. 개체 연결 원시 말뭉치의 경우 위 작업을 시행한 뒤 작업 대상인 5개 대분류 (PER, AFW, ORG, LOC, DAT)만 선택하여 추출한다. 본 작업을 통해 작업자들의 편의성을 증진하여 구축 작업이 원활히 이루어질 수 있도록 준비한다. 다음으로 구축 작업에서는 먼저 언어 분석기(자동화 시스템)의 결과를 얻어 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 구축 작업자가 수정하는 작업을 거친다. 변경된 구축 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 조회

하고 이에 따라 작업자에게 줄 작업 할당량을 정하여 자동 또는 수동으로 작업물을 부여한다. 또한 작업자가 추진 일정 속도에 맞추어 구축 작업을 잘 이행하고 있는지 관리한다.

마지막으로 지침 관리 과정에서는 작업 진행 중 지침 변경사항을 버전별로 관리하고 작업자들에게 공지하여 재교육한다.

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

개체명 부착 작업자 작업방식		
도구	개체명 부착 도구 Kane	클라우드 소싱 플랫폼 Wisedome
작업자	개체명 작업자	개체명 작업자
분량	1회 약 40만 어절 배포, 작업 기간 일주일	1회 약 80만 어절 배포, 작업 기간 일주일
작업 대상	문어/구어 원시 말뭉치 300만 어절	개체명 대분류 15개를 기준으로 표지가 부착된 '19 개체명 말뭉치 300만 어절
방식	<ul style="list-style-type: none"> 같은 군집화(유사한 개체명이 포함된 문장)에 포함된 문장을 여러 작업자가 확인 품질을 평가할 때 다면 평가를 진행 	<ul style="list-style-type: none"> 기계로 1차 확장작업(15개에서 150개 세분류로 수정하는 작업)을 거친 말뭉치를 작업자가 확인 각 작업자에 담당할 개체명 범주(3개)를 부여하고 교육 및 작업을 진행, 각 문장은 여러 작업자의 결과를 기반으로 다면 평가 수행 다면 평가에서 결과가 일치하지 않은 개체명은 다시 플랫폼에 업로드하여 재작업을 진행하고 일치 여부를 추가 확인
평가 시기	<ul style="list-style-type: none"> 한 달에 한 번 평가 말뭉치와 비교하여 수정 작업 진행 누적 작업량이 40%가 넘는 시점부터 매주 자동화 시스템을 통해 도출된 오류 사항에 대한 수정 작업이 이루어짐. 	<ul style="list-style-type: none"> 제출된 작업물을 대상으로 일관성 검사 진행, 도출된 오류 사항에 대한 수정 작업이 이루어짐

<표 6> 개체명 부착 작업자 작업방식

평가 작업자 작업방식	
도구	개체명 부착 도구 Kane, 클라우드 환경 도구 활용
분량	<ul style="list-style-type: none"> 문어/구어 원시 말뭉치 300만 어절 중 10%를 선별 평가용으로 구축 주석 작업 평가를 위하여 초벌 태깅은 한 달 내에 수행
방식	<ul style="list-style-type: none"> 작업자가 전문가이므로 다면 평가가 아닌 구축 도구를 이용하여 작업 비정기적 작업: 변경된 지침을 반영하여 평가 말뭉치 수정 수정 작업: 기구축된 개체명의 오류를 찾아내어 말뭉치 수정

<표 7> 평가 작업자 작업방식

개체 연결 작업자 작업방식	
도구	개체명 부착 도구 Kane
작업 대상	'19 개체명 말뚝치 300만 어절, '20 개체명 말뚝치 500만 어절, '21 개체명 말뚝치 300만 어절
분량	연도별로 나누어 배포, 작업 기간 약 15일
방식	<ul style="list-style-type: none"> 제출된 작업물을 대상으로 일관성 검사 진행, 도출된 오류 사항에 대한 수정 작업이 이루어짐. 개체 연결이 잘못된 경우(페이지가 없는 경우 등)는 재작업을 수행하고, 동일 개체명에 다른 위키피디아 페이지가 연결된 경우 또는 지침과 일치하지 않는 경우가 발견되면 유사한 개체명을 추출하여 추가 작업을 요청하는 방식으로 진행

<표 8> 개체명 연결 작업자 작업방식

또한 한정된 사업 기간 내에 고품질의 말뚝치를 완성해내기 위해서 구체적으로 일정을 계획하였고 구축 작업은 지연 없이 잘 진행되었다. 실제 구축 명세는 다음과 같다.

21년 신규 말뚝치 구축 및 수정				
차수	일시	작업 명	작업량 (단위:문장)	누적 작업량 (단위:문장)
1	6월 마지막 주 (21/06/28 ~ 21/07/13)	교육1	33,000	-
2	7월 셋째 주 (21/07/08 ~ 21/07/19)	교육1 -피드백	33,000	-
3	7월 넷째 주 (21/07/20 ~ 21/07/26)	작업1	44,000	44,000
4	7월 마지막 주 (21/07/27 ~ 21/08/02)	작업2	44,000	88,000
5	8월 첫째 주 (21/08/03 ~ 21/08/09)	작업3	44,000	132,000
6	8월 둘째 주 (21/08/10 ~ 21/08/18)	국어원 평가 시스템 작업		
7	8월 셋째 주 (21/08/19 ~ 21/08/25)	작업4	44,000	176,000
8	8월 넷째 주 (21/08/26 ~ 21/09/01)	작업5	44,000	220,000
9	9월 첫째 주 (21/09/02 ~ 21/09/08)	작업6	44,000	264,000
10	9월 둘째 주 (21/09/09 ~ 21/09/15)	작업7	36,641	300,641
전체		300,641		

11	9월 셋째 주 (21/09/27 ~ 21/10/03)	수정 작업1	22,482	-
12	10월 첫째 주 (21/10/05 ~ 21/10/11)	수정 작업2	23,286	45,768
13	10월 둘째 주 (21/10/12 ~ 21/10/18)	수정 작업3	22,671	68,439
14	10월 셋째 주 (21/10/19 ~ 21/10/25)	수정 작업4	27,782	96,221
15	10월 넷째 주 (21/10/26 ~ 21/11/01)	수정 작업5	26,176	122,397
16	11월 첫째 주 (21/11/02 ~ 21/11/08)	수정 작업6	35,365	157,762
17	11월 둘째 주 (21/11/09 ~ 21/11/15)	수정 작업7	29,714	187,476
18	11월 셋째 주 (21/11/16 ~ 21/11/22)	〈우리말샘〉 사전 비교 작업		
19	11월 넷째 주 (21/11/23 ~ 21/11/29)	〈우리말샘〉 사전 비교 작업		

<표 9> '21 신규 말뭉치 구축 및 수정 작업 내용

19년 개체명 말뭉치 확장				
차수	일시	작업 명	작업량 (단위:문장)	누적 작업량 (단위:문장)
1	8월 넷째 주 (21/08/23 ~ 21/08/31)	작업1	111,460	111,460
2	9월 첫째 주 (21/09/01 ~ 21/09/09)	작업2	111,460	222,920
3	9월 둘째 주 (21/09/10 ~ 21/09/18)	작업3	111,460	334,380
4	9월 넷째 주 (21/09/23 ~ 21/09/30)	작업4	111,386	445,766
5	10월 첫째 주 (21/10/06 ~ 21/10/14)	작업5	104,667	550,433
전체		550,433		

<표 10> '19 개체명 말뭉치 확장작업 내용

개체 연결 구축			
차수	일시	작업 명	작업 내용
1	21/07/12 ~ 21/08/19	작업1	2020년 개체명 분석 말뭉치 개체 연결 작업 및 지침 정리
2	21/09/09 ~ 21/09/24	작업1	2020년 개체명 분석 말뭉치 개체 연결 작업 검수 1차
3	21/09/27 ~ 21/10/11	작업1	2020년 개체명 분석 말뭉치 개체 연결 작업 검수 2차
4	21/10/18 ~ 21/10/30	작업2	2020년 개체명 분석 말뭉치 개체 연결 작업 중 중의성 해소되지 않은 문장 수정
5	21/11/02 ~ 21/11/16	작업3	2019년 개체명 분석 말뭉치 개체 연결 작업
6	21/11/17 ~ 21/11/28	작업4	2021년 개체명 분석 말뭉치 개체 연결 작업
7	21/11/29 ~ 21/12/10	작업5	전체 말뭉치 중 미완료 문장 작업

<표 11> 개체 연결 구축 작업 내용

21년 평가 말뭉치 구축 및 수정				
차수	일시	작업 명	작업량 (단위:문장)	누적 작업량 (단위:문장)
1	6월 마지막 주 (21/06/28 ~ 21/07/05)	수정 작업1	15,000	15,000
2	7월 둘째 주 (21/07/06 ~ 21/07/13)	수정 작업2	15,000	30,000
3	7월 셋째 주 (21/07/14 ~ 21/07/20)	수정 작업3	15,000	45,000
4	7월 넷째 주 (21/07/21 ~ 21/07/27)	수정 작업4	15,000	60,000
5	7월 마지막 주 (21/07/28 ~ 21/08/03)	수정 작업5	15,000	75,000
6	8월 첫째 주 (21/08/04 ~ 21/08/09)	수정 작업6	9,644	84,644
7	8월 둘째 주 (21/08/10 ~ 21/08/18)	국어원 평가 시스템 작업		
8	8월 셋째 주 (21/08/19 ~ 21/08/25)	평가 작업1	12,000	12,000
9	8월 넷째 주 (21/08/26 ~ 21/09/01)	평가 작업2	12,000	24,000
10	9월 첫째 주 (21/09/02 ~ 21/09/08)	평가 작업3	12,000	36,000
11	9월 둘째 주 (21/09/09 ~ 21/09/15)	평가 작업4	5,000	41,000
전체		수정 84,644문장, 평가 41,000문장		

<표 12> '21 평가 말뭉치 구축 및 수정 작업 내용

3.2. 말뭉치 납품

국립국어원으로 전달한 말뭉치 납품 명세는 다음과 같다. 전달한 말뭉치 파일명은 편의를 위해 개체명 말뭉치가 구축된 연도를 기준으로 표기하였다.

말뭉치 납품 결과	
일시	전달품목
21.09.27	<ul style="list-style-type: none"> • '20 개체 연결 분석 말뭉치 • '21 개체명 분석 말뭉치
21.11.15	<ul style="list-style-type: none"> • '19 개체명 분석 말뭉치
21.11.19	<ul style="list-style-type: none"> • '19 개체명 분석 말뭉치 1차 수정본 • '20 개체명 분석 말뭉치 1차 수정본 • '20 개체 연결 분석 말뭉치 • '21 개체명 분석 말뭉치 1차 수정본 • '21 개체 연결 분석 말뭉치
21.12.07	<ul style="list-style-type: none"> • '19 개체명 분석 말뭉치 2차 수정본 • '20 개체명 분석 말뭉치 2차 수정본 • '21 개체명 분석 말뭉치 1차 수정본
21.12.09	<ul style="list-style-type: none"> • '19 개체 연결 분석 말뭉치 • '20 개체 연결 분석 말뭉치 1차 수정본 • '21 개체 연결 분석 말뭉치 1차 수정본
21.12.13	<ul style="list-style-type: none"> • '20 개체 연결 분석 말뭉치

<표 13> 말뭉치 납품 결과

4. 검증 및 산출물 보고

4.1. 검증 절차

본 사업에서는 2021년 개체명 말뭉치에 대한 검증이 수행되어야 한다. 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 다면 평가를 진행한다. 군집화는 유사한 개체명이 있는 문장들로 구성되어 있으므로 군집화에 포함된 문장을 다수의 작업자가 작업하여 문장에 포함된 개체명의 범주가 일치하는지를 확인하는 것이다. 다면 평가 검증은 작업자가 확인이 필요한 문장을 선택하기 위해 사용되고, 약 한 달에 한 번씩 작업자에게 피드백을 제공한다.

다음은 평가 말뭉치와의 비교 평가이다. 2021년 개체명 말뭉치 전체의 10%(약 4만 1천 문장)에 해당하는 말뭉치를 검증팀과 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 검증팀의 결과를 answer, 성능을 비교할 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 정밀도로 result 데이터 개체명 중 기준인 answer와 일치하는 비율을 말한다. Recall은 재현율로 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 다면 평가를 통해 수행됐다. 예측 평가는 최종 말뭉치 검증 작업에서 수행되는 것으로 다면 평가와 유사하다. 다면 평가가 해당 작업 회차에 대한 비교라면 예측 평가는 전체 말뭉치를 기준으로 수행한다. 수행하여 얻어진 비교 결과물은 최종 말뭉치 수정에 사용하였다.

4.2. 검증 결과

□ 평가 말뭉치와의 비교 평가

위에서 설명한 두 번째 검증 방식으로 2021년 개체명 말뭉치를 검증하였고, <표 14>는 검증 결과이다. 평가 말뭉치로 구축된 41,000문장에 대해 검증팀과 구축팀 결과를 비교하였다.

품질 측정 결과	
Precision	96.8
Recall	97.52
F1	97.16

<표 14> 품질 측정 결과

4.3. 개체명 사전

개체명 사전은 개체명을 중심으로 개체명 범주 및 개체 연결에서 얻어진 정보를 통합한 결과물이다. 개체명 사전은 JSON 형식의 텍스트로 구조는 아래와 같다.


```

{
  "id": "0000026672",
  "form": "홍익대",
  "ne": [
    {
      "label": "LC_OTHERS",
      "kid": "05120000009781",
      "wikiform": "홍대 (지명)",
      "definition": "NA",
      "wikiLink": "https://ko.wikipedia.org/wiki/%ED%99%8D%EB%8C%80_(%EC%A7%80%EB%AA%85)",
      "SynonymGroup": [ "Hongdae", "honda", "hongikuniversity", "honggikuniversity", "hongdae", "홍대", "홍
머" ],
      "examples": [
        [ "NWRW1800000038.288.4.3", "홍익대 앞 한국미술정보센터 김달진 소장이다." ]
      ]
    },
    {
      "label": "OGG_EDUCATION",
      "kid": "04010000003074",
      "wikiform": "홍익대학교",
      "definition": "대한민국 서울특별시와 세종특별자치시의 사립 종합대학",
      "wikiLink": "https://ko.wikipedia.org/wiki/%ED%99%8D%EC%9D%B5%EB%8C%80%ED%95%99%E
A%B5%90",
      "SynonymGroup": [ "홍익대학교", "홍익유니버시티", "홍대", "홍머" ],
      "examples": [
        [ "NPRW2000000001.30736.2.4", "화재 사고 조사위원회를 이끈 김정훈 홍익대 전기전자공학부 교수와 최
윤석 울산과학기술원(UNIST) 에너지공학부 교수가 기술적인 설명을 많이 했지요." ]
      ]
    }
  ]
}

```

<표 15> 개체명 사전 구조 예제

예제와 같이 ‘홍익대’ 개체명을 기준으로 정보를 표현하고 있다. ‘id’는 전체 개체명에서 ‘홍익대’ 개체명의 순번을 의미하고 ‘ne’ 구조는 ‘홍익대’ 개체명이 가질 수 있는 개체명 정보를 표현한다. 첫 번째 개체명 범주는 LC_OTHERS이고 두 번째는 OGG_EDUCATION이다. 각 개체명 범주 내에는 다음의 정보를 가지고 있다.

- Label : 개체명 150개 범주를 뜻한다.
- kid : 개체 연결 결과물에 사용되는 id로 k는 knowledge를 줄인 표현이다. kid 구성은 14개의 숫자로 구성된다. 14개 중 왼쪽에서 처음 두 자리는 개체명 상위 범주 15개를 의미하고 다음 두 자리는 각 상위 범주에 속한 하위범주의 번호이다. 나머지 열 자리는 각 하위범주에 속하는 개체명을 정렬하여 번호를 부여한다.
- wikiform : 위키피디아 페이지에 연결된 경우 해당 페이지의 제목을 나타낸다.

- definition : 위키피디아에서 정의한 정보를 나타낸다.
- wikiLink : 연결된 위키피디아 페이지를 나타낸다.
- SynonymGroup : 동일한 위키피디아에 연결된 개체명으로 유의어를 나타낸다.
- examples : 해당 개체명이 나타난 문장의 ID와 문장 최대 5개를 표시한다.

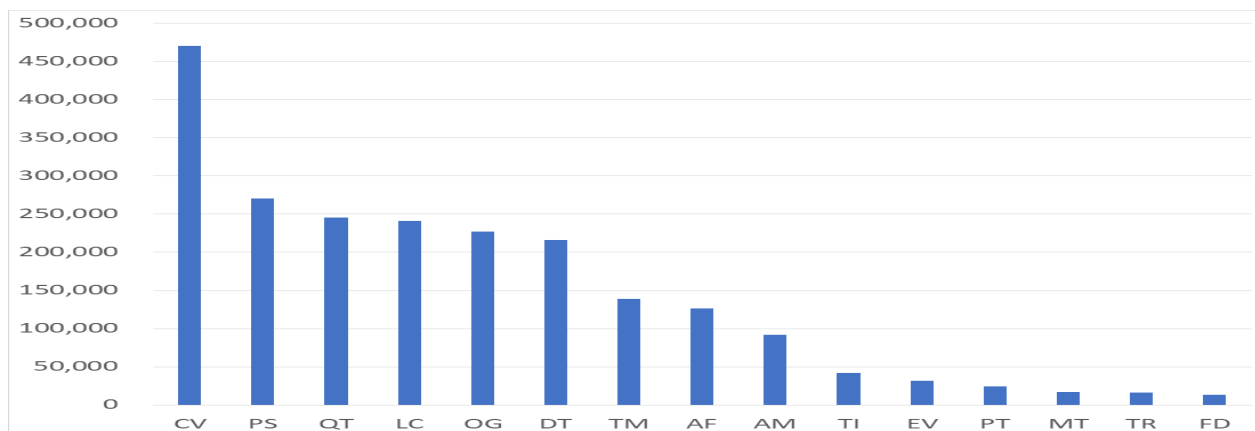
□ 개체명 말뭉치 분석 결과

```

{
  "id": "NXNE2107020",
  "metadata": {
    "title": "국립국어원 신문 말뭉치 추
출 NXNE2107020",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2019",
    "category": "신문 > 전국종합지",
    "annotation_level": "개체명 분석",
    "sampling": "본문 전체"
  },
}

"document": [
{
  "id": "NIRW2000000001.91",
  "metadata": {
    "title": "노컷뉴스 2019년 기사",
    "author": "CBS노컷뉴스 광영식 기자",
    "publisher": "노컷뉴스",
    "date": "20190102",
    "topic": "IT/과학",
    "original_topic": "IT_과학,보안IT_과학,모바일IT_과학,인터넷_SNS"
  },
  "sentence": [
    {
      "id": "NIRW2000000001.91.1.1",
      "form": "종합적인 해양수산정보 관리 기반 마련...121억 투입"
    }
  ]
}
]

```



전체 개체명 분석 구축 결과 대분류별 출현 빈도는 <그림 15>와 같다.

주로 의식주, 사회, 예술, 문화, 제도, 법과 관련된 CIVILIZATION이 470,096개로 다른 대분류에 비해 월등히 많이 부착되었다. CIVILIZATION은 가장 많은 21개의 세부 분류를 가지고 있다.

다음으로는 인명, 별명, 캐릭터 명, 애완동물 이름과 관련된 PERSON이 270,291개 부착되었다.

다음으로는 수량과 순서 등 기타 수사와 관련된 QUANTITY가 245,184개이고 지역, 장소, 지형, 지리 등과 관련된 LOCATION이 241,161개 부착되었다.

이외에 기관과 단체를 나타내는 ORGANIZATION이 227,124개, 날짜를 나타내는 DATE 215,827개, 14개의 대분류에 포함되지 않는 나머지 개체명 대상인 색, 세포, 장치, 질병, IT 서비스 등을 포함하는 TERM은 138,818개, 인공물과 관련된 ARTIFACTS는 126,270개, 동물과 신체 및 몸의 명칭과 관련된 ANIMAL은 92,136개가 나타났다. 그 외 시간을 나타내는 TIME은 41,862개, 사건 및 행사 등과 관련된 EVENT는 31,420개, 식물과 관련된 PLANT가 24,576개가 나타났으며 원소 및 화학물을 나타내는 MATERIAL은 16,757개, 특정 이론, 법칙, 원리 등과 관련된 THEORY는 16,297개, 학문 분야를 나타내는 STUDY_FIELD는 13,010개로 가장 적게 부착되었다. 2020년 개체명 말뭉치 분석 결과와 비교하면 본 사업에서는 문장의 수가 2배로 증가하였고 상위 6개의 개체명 범주의 출현 빈도는 그에 따라 증가하였다. 하지만 나머지 9개의 개체명 범주는 빈도가 비슷하고 증가세가 높지 않다. 상위 6개는 고유명사와 같이 시간에 따라 증가하는 항목들로 구성되고 그 외 범주는 고정적인 항목이 많다는 것을 유추할 수 있다.

각 대분류에 속하는 세부 분류별 출현 빈도는 아래의 <표 17>과 같다.

CIVILIZATION	빈도
CV_FOOD	184633
CV_POSITION	68237
CV_OCCUPATION	60274
CV_RELATION	53454
CV_DRINK	40486
CV_CLOTHING	13430
CV_ART	10934
CV_SPORTS	7719
CV_LAW	6730
CV_LANGUAGE	4915
CV_TRIBE	4683
CV_PRIZE	3296
CV_FOOD_STYLE	2800
CV_POLICY	2567
CV_SPORTS_POSITION	1642
CV_SPORTS_INST	1131
CV_TAX	901
CV_FUNDS	838
CV_CULTURE	595
CV_CURRENCY	556
CV_BUILDING_TYPE	275

DATE	빈도
DT_DAY	73965
DT_DURATION	46591
DT_YEAR	35631
DT_OTHERS	32099
DT_SEASON	12918
DT_MONTH	9453
DT_WEEK	3089
DT_DYNASTY	1788
DT_GEOAGE	293

THEORY	빈도
TR_SCIENCE	8776
TR_MEDICINE	3465
TR_SOCIAL_SCIENCE	2410
TR_ART	845
TR_HUMANITIES	533
TR_OTHERS	268

PERSON	빈도
PS_NAME	236561
PS_CHARACTER	23822
PS_PET	9908

LOCATION	빈도
LCP_COUNTRY	64919
LCP_CITY	43582
LC_OTHERS	37790
LCP_COUNTY	36908
LCP_PROVINCE	17739
LCP_CAPITALCITY	15062
LC_SPACE	6582
LCG_CONTINENT	5385
LCG_MOUNTAIN	4265
LCG_ISLAND	3845
LCG_RIVER	3087
LCG_OCEAN	1048
LCG_BAY	949

QUANTITY	빈도
QT_COUNT	76166
QT_ORDER	45660
QT_MAN_COUNT	27067
QT_PRICE	23747
QT_PERCENTAGE	21247
QT_AGE	16818
QT_OTHERS	12003
QT_SPORTS	6034
QT_LENGTH	5867
QT_WEIGHT	3810
QT_ADDRESS	1907
QT_SIZE	1327
QT_VOLUME	1171
QT_TEMPERATURE	959
QT_PHONE	705
QT_SPEED	364
QT_ALBUM	277
QT_CHANNEL	55

ORGANIZATION	빈도
OGG_ECONOMY	63085
OGG_FOOD	53780
OGG_POLITICS	47848
OGG_SPORTS	11900
OGG_EDUCATION	11469
OGG_OTHERS	10527
OGG_MEDIA	7193
OGG_ART	5847
OGG_MILITARY	4280
OGG_HOTEL	2831
OGG_MEDICINE	2687
OGG_LAW	2626
OGG_RELIGION	1969
OGG_SCIENCE	726
OGG_LIBRARY	356

ARTIFACTS	빈도
AFA_VIDEO	25369
AFW_OTHER_PRODUCTS	24303
AF_TRANSPORT	21391
AFA_DOCUMENT	11984
AF_BUILDING	10271
AFA_MUSIC	8206
AF_ROAD	6777
AF_CULTURAL_ASSET	5194
AF_WEAPON	4483
AFA_PERFORMANCE	3222
AF_MUSICAL_INSTRUMENT	3064
AFA_ART_CRAFT	1122
AFW_SERVICE_PRODUCTS	884

TERM	빈도
TMI_HW	30002
TMI_SERVICE	21665
TMM_DISEASE	21567
TM_DIRECTION	19017
TM_CELL_TISSUE_ORGAN	14664
TM_COLOR	10922
TMI_SITE	4340
TM_SPORTS	4311
TMM_DRUG	3757
TMI_SW	2855
TM_SHAPE	2456
TMI_PROJECT	1710
TMIG_GENRE	940
TMI_MODEL	501
TMI_EMAIL	79
TM_CLIMATE	32

ANIMAL	빈도
AM_PART	59696
AM_MAMMALIA	18591
AM_OTHERS	3980
AM_BIRD	2317
AM_INSECT	2220
AM_REPTILIA	1979
AM_TYPE	1535
AM_FISH	1451
AM_AMPHIBIA	367

TIME	빈도
TI_DURATION	29737
TI_OTHERS	6384
TI_HOUR	5400
TI_MINUTE	253
TI_SECOND	88

PLANT	빈도
PT_FRUIT	5688
PT_GRASS	5163
PT_PART	5038
PT_FLOWER	4697
PT_TREE	1829
PT_OTHERS	1304
PT_TYPE	857

EVENT	빈도
EV_OTHERS	12703
EV_FESTIVAL	8468
EV_SPORTS	8192
EV_WAR_REVOLUTION	1370
EV_ACTIVITY	687

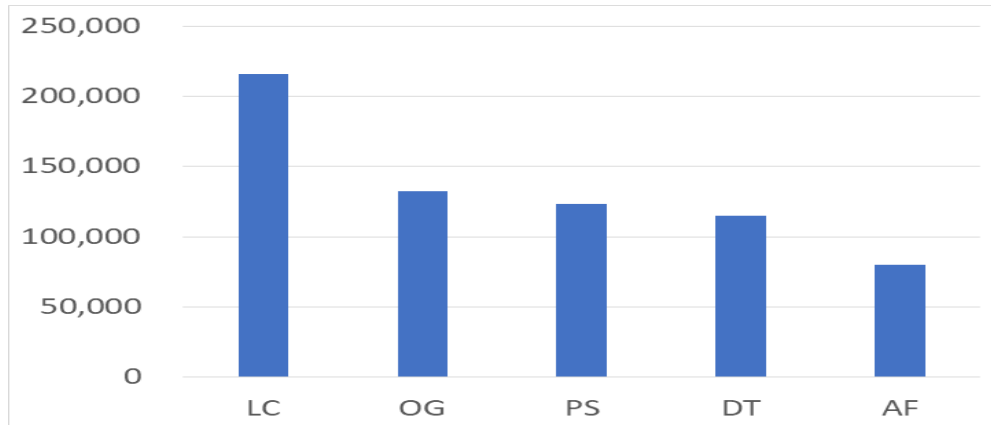
MATERIAL	빈도
MT_CHEMICAL	12014
MT_ELEMENT	2473
MT_METAL	1465
MT_ROCK	805

STUDY_FIELD	빈도
FD_SCIENCE	6535
FD_HUMANITIES	2686
FD_MEDICINE	2150
FD_SOCIAL_SCIENCE	1043
FD_ART	474
FD_OTHERS	122

<표 17> 세부 분류 전체 빈도

□ 개체 연결 말뭉치 분석 결과

본 사업의 개체 연결 말뭉치는 개체명 원시 말뭉치와 동일하다. 구조는 위에서 설명한 JSON 구조와 동일하다. 아래 표는 15개의 상위 개체명 범주 중 개체 연결의 대상이 되는 5가지 범주에 대해 위키피디아로 개체 연결 주석 작업이 된 개체명의 빈도를 표현한다.



<그림 16> 위키페이지가 연결된 빈도

가장 연결이 많이 이루어진 범주는 지역, 장소에 해당하는 LOCATION으로 전체 개체명이 241,161개이고 개체 연결이 된 항목은 216,215개이다. 약 90%의 LOCATION 개체명이 위키피디아 페이지로 연결됨을 확인할 수 있다. 다음은 ORGANIZATION으로 전체 227,124개 개체명에서 132,582개의 개체명이 연결되었다. 절반을 넘는 수준으로 아직 위키피디아에 다양한 조직에 대한 정보가 없는 것을 확인할 수 있다.

다음 PERSON 범주는 270,291개의 개체명 중 123,737개가 연결되었고 애견과 관련된 PS_PET, 위키피디아에 존재하지 않는 인명이 문서에 많음을 확인하였다. 유사하게 DATE도 215,827개 개체명 중 114,864개 개체명이 연결되었고 이것은 기간에 해당하는 DT_DURATION이 연결되지 않는 항목이 많았고 연월일이 있는 명확한 날짜를 표현한 개체명이 많음에 비해 위키피디아 페이지에는 이벤트가 있는 날 위주로 등록되어 있어 해당 개체명이 연결되지 않아 빈도가 낮은 것으로 분석된다. 마지막으로 ARTIFACTS는 126,270개 중 80,285개로 연결 대상 개체명은 적지만, 개체 연결 비율은 높았다. 각 대분류에 속하는 세부 분류별 출현 빈도는 아래 <표 18>과 같다.

ORGANIZATION	빈도
OGG_POLITICS	39885
OGG_ECONOMY	37741
OGG_FOOD	11191
OGG_SPORTS	10180
OGG_EDUCATION	9350
OGG_MEDIA	6409
OGG_OTHERS	4175
OGG_MILITARY	3710
OGG_ART	3320
OGG_LAW	2433
OGG_RELIGION	1420
OGG_MEDICINE	1325
OGG_HOTEL	862
OGG_SCIENCE	393
OGG_LIBRARY	188

ARTIFACTS	빈도
AFA_VIDEO	20761
AF_TRANSPORT	19126
AFW_OTHER_PRODUCTS	10629
AF_BUILDING	5409
AFA_DOCUMENT	4569
AF_CULTURAL_ASSET	4197
AF_ROAD	4153
AF_WEAPON	4150
AF_MUSICAL_INSTRUMENT	2966
AFA_MUSIC	2519
AFA_PERFORMANCE	1297
AFA_ART_CRAFT	336
AFW_SERVICE_PRODUCTS	173

LOCATION	빈도
LCP_COUNTRY	64709
LCP_CITY	37101
LCP_COUNTY	32734
LC_OTHERS	26132
LCP_PROVINCE	17483
LCP_CAPITALCITY	15015
LC_SPACE	6421
LCG_CONTINENT	5354
LCG_ISLAND	3505
LCG_MOUNTAIN	3329
LCG_RIVER	2507
LCG_OCEAN	1008
LCG_BAY	917

DATE	빈도
DT_DAY	53160
DT_YEAR	21775
DT_DURATION	15254
DT_SEASON	11764
DT_MONTH	5795
DT_OTHERS	5431
DT_DYNASTY	1432
DT_GEOAGE	253

PERSON	빈도
PS_NAME	118321
PS_CHARACTER	5385
PS_PET	31

<표 18> 세부 분류에 연결된 위키페이지 전체 빈도

위키피디아를 기준으로 가장 많이 연결된 페이지 상위 10개를 확인하였다. 가장 많이 연결된 페이지는 ‘오늘’로 DT_DAY의 절반이 ‘오늘’에 연결되었음을 알 수 있다. 그리고 뉴스 등의 문서에서 국가에 관련된 문장이 많아 ‘대한민국’, ‘미국’, ‘일본’, ‘제주’, ‘중화인민공화국’, ‘조선민주주의인민공화국’이 다수 출현하였고 개체 연결이 많이 이루어졌음을 알 수 있다.

위키피디아 제목	빈도
오늘	26959
대한민국	13013
미국	9095
서울특별시	8893
일본	7500
정부	7044
제주특별자치도	6957
날	6003
중화인민공화국	5962
조선민주주의인민공화국	5148

<표 19> 가장 많이 연결된 위키피디아 페이지(제목)

4.5. 사업 보고

국립국어원과 사업의 진행 척도와 내용을 공유하기 위하여 진행한 보고는 다음과 같다.

☐ 착수 보고

연구 수행 계획서에 사업 수행 목표 및 범위를 명시하였으며 추진 전략과 과업별 수행 방안을 설명하였다. 수행 조직별 참여 인력과 역할을 기술하였으며, 구체적인 수행 일정과 총사업비 총괄표를 첨부하였다. 착수계 및 선금 신청서 및 각서 등 필요 서류를 제출하였으며 화상 회의를 통해 보고를 진행하였다.

☐ 주간 보고

매주 월요일마다 사업이 진행된 내용을 기술팀과 언어팀에서 각각 기술하여 보고서로 제출하였다.

☐ 월간 보고

매달 17일마다 사업이 진행된 내용을 기술팀과 언어팀에서 각각 기술하여 보고서로 제출하였다.

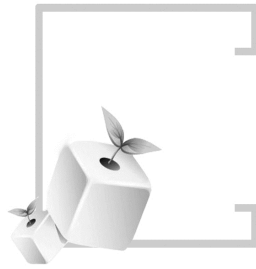
☐ 중간보고

사업 진행 중간 점검을 위해 현재까지 구축된 말뭉치의 양과 지침 구축 및

수정사항 등을 공유하고, 지침 재정비 과정에서 발생한 문제점을 해결하기 위해 화상 회의를 통해 보고를 진행하였다.

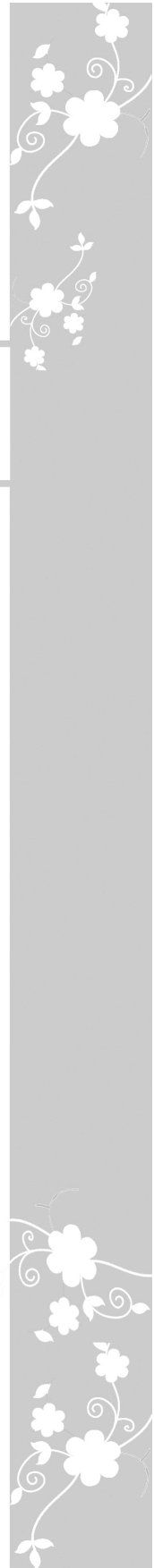
☐ 최종 보고

사업 최종 보고를 위해 완료계와 용역 보고서를 제출하고, 전체 사업 진행 내용에 대한 사항과 사업 결과에 대한 보고를 진행하였다.



제 3 장

개체명 및 개체 연결 지침



1. 개체명 말뭉치 구축 지침 정비

1.1. 지침 주요 수정사항

본 사업에서는 ‘2020년 개체명 말뭉치 연구 분석’ 최종보고서에 포함된 <2020년 개체명 분석 말뭉치 구축 지침>을 정비하여 <2021년 개체명 분석 말뭉치 구축 지침>을 작성하였다.

올해 사업은 <2020년 개체명 분석 말뭉치 구축 지침>의 체계화 및 정밀화, 일관성 향상을 목표로 하고 있다. 작업자들이 분석 작업 시 발견되는 문제점에 대하여 문의 사항을 적으면 분석 담당자가 그에 대한 답을 하였다. 문의 사항 중에 논의가 필요한 내용이 있으면 연구진 내의 논의를 거쳐 국립국어원과 협의하여 지침을 수정하는 순환적인 작업을 사업 기간 내내 수행하였다. <2021년 개체명 분석 말뭉치 구축 지침>은 2020년과 마찬가지로 15개의 대분류와 145개의 소분류, 7개의 세분류로 구성되어 있다. 즉, 개체명 태그의 명칭이 신설되거나 수정된 내용은 없다. 이러한 순환적인 작업을 통해 지침 내용을 보다 체계화하고 정밀화하였다. 이에 대한 수정 내용은 ‘1.2. 개체명 분석의 대원칙 정비’와 ‘1.3. 세부 분류별 수정사항’에 기술하도록 하겠다.

2020년 사업에서는 대분류 담당자가 각각 하나의 대분류를 맡아 지침을 수정하였다. 최말단 태그의 모든 구성은 ‘정의 및 설명, 분석 대상, 예문, 주의 사항’의 순으로 기술하였다. 전체적인 구성은 같지만, 대분류 아래의 설명 유무, 분석 대상의 기술 방식, 예문의 제시 방식, 주의 사항의 설명 방식 등에서 상이한 부분들이 발견되었다. 올해 사업에서는 15개의 대분류의 설명 방식의 일관성을 확보하였다. 첫째, 대분류의 정의를 일관성이 있게 기술하였다. 또한 해당 대분류에 속하는 소분류 및 세분류를 한눈에 확인할 수 있도록 분류 명칭과 정의 및 설명을 표로 간단히 제시하였다. 둘째, 대분류 주의 사항을 추가하였다. 이는 소분류 및 세분류를 확인하기 전에 해당 대분류에서 특기할 사항을 정리한 내용이다. 셋째, 각 소분류 및 세분류의 정의 및 설명, 분석 대상 중에 상하 관계가 일관되게 정리되지 못한 부분을 상하 관계가 명확히 드러나도록 기술하였다. 넷째, 각 소분류 및 세분류의 주의 사항 부분을 대원칙의 순서에 맞게 ‘태깅 대

상’, ‘태깅 단위’, ‘태깅 범주’의 순서에 맞게 기술하였다. 또한 무엇에 대한 설명인지 알 수 있게 개체명 또는 태그를 앞부분에 제시하였다. 다섯째, 2020년 지침에 주의 사항으로 기술된 것 중 예외적인 내용이 아닌 것은 분석 대상 항목으로 이동하여 기술하였다.

이상의 내용을 1.7. 버전부터 2.1. 버전까지 총 다섯 번에 걸쳐 수정하였다.

1.2. 개체명 분석의 대원칙 정비

개체명 분석을 하기 위해서는 태깅 대상(무엇을 분석할 것인가), 태깅 단위(어디서부터 어디까지를 분석할 것인가), 태깅 범주(어떤 태그를 부착할 것인가)의 세 가지 차원에서 결정이 이루어져야 하는데 대원칙에는 이에 대한 설명이 제시되어 있다. 그밖에 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 대한 지침 방법을 제시하고 있다. 이와 같은 대원칙의 내용은 2020년 지침 내용을 유지하고 있다. 2021년 작업에서는 <우리말샘> 사전 등재 여부, 등재 방식이 대원칙의 내용에 추가되었다.

첫째, 태깅 대상과 관련하여 <우리말샘>을 참고한다. <우리말샘> 사전에 등재되어 있는 보통명사라면 150개 항목의 태그로 태깅할 수 있는 개체명 대상 후보가 된다.

둘째, 태깅 단위와 관련하여 <우리말샘>을 참고한다. <우리말샘> 사전에 어떻게 등재되어 있는지를 통해 태깅 단위를 결정할 수 있다. <우리말샘> 사전 검색 결과 해당 명사 구성이 ‘한 단어’로서 등재되어 있는 경우(‘-’, ‘^’와 같은 기호와 함께 등재되어 있는 경우) ‘최장 단위 태깅’을 하여 전체를 하나로 태깅한다. <우리말샘>에 ‘구’로서 등재되어 있거나(띄어쓰기가 되어 있는 경우), 등재되어 있지 않은 경우에도 해당 구가 본래의 의미와 달라지면 ‘최장 단위 태깅’을 하여 전체를 하나로 태깅한다. 이외에는 ‘최소 단위 태깅’을 하여 각 단어를 따로 태깅한다.

예를 들어, ‘힙합 댄서’, ‘상승 가수’는 <우리말샘> 사전에 ‘힙합 댄

서’, ‘상송 가수’와 같이 ‘구’로 등재되어 있다. 따라서 ‘최장 단위 태깅’을 하여 전체를 하나로 태깅한다. 2020년 작업에서는 ‘최장 단위 결정의 원리-보통명사’에 ‘사촌 동생’과 같은 구 구성의 경우는 최장 단위 태깅을 하였는데, ‘직업’의 경우에는 구 구성을 최소 단위 태깅을 하였다. 2021년 작업에서는 ‘직업’, ‘직위’, ‘가족’의 경우 <우리말샘> 사전 등재 여부에 따라 최장 단위 태깅하여 지침에 일관성을 확보하였다.

또한 예외적으로 복합어 내부 요소를 태깅하는 경우에도 <우리말샘> 사전의 등재 방식을 통해 결정할 수 있다. 원소명, 금속물 등 MT는 해당 원소명 및 금속물 등의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다. <우리말샘> 사전에 ‘^’ 기호를 사용하여 등재되어 있는 경우, MT에 한해 ‘^’를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)하기로 하였다.

1.3. 세부 분류별 분석 지침 정비

□ 회의 및 협의 내용

지침의 완성도를 높이기 위하여 여러 차례 협의 및 회의가 이루어졌다.

차수	참여 기관	협의 내용
1차	국립국어원 데이터리 한림대학교	1. CV_POSITION <ul style="list-style-type: none"> ‘기관+직위’ 태깅 단위 문제 ‘삼성전자 사장’과 같이 분리되는 경우 ‘기관’은 OGG 계열로, ‘직위’는 CV_POSITION으로 태깅함. 예) 삼성전자/OGG_ECONOMY 사장/CV_POSITION 단, ‘선거대책위원장’과 같이 기관과 직위가 분리되지 않는 경우는 전체를 CV_POSITION으로 태깅함. ‘부서+직위’ 태깅 단위 문제 부서+직위의 형태로 이루어져 있으나, 전체가 정식 직위 명칭이거나 분리하기 어려운 경우 전체를 태깅함. 따라서 ‘언어정보과장’은 전체를 CV_POSITION으로 태깅함.
2차	국립국어원 데이터리 한림대학교	1. MT 관련 (복합어 내부 태깅) <ul style="list-style-type: none"> 2020년 지침의 내용은 다음과 같다. MT는 해당 원소명 및 금속물 등 물질의 의미를 유지하고 있을 때 복합어

	경희대학교	<p>내부 요소임에도 태깅할 수 있다. 따라서 ‘산소마스크’, ‘마그네슘 리본’에서 ‘산소’와 ‘마그네슘’은 MT로 태깅하였다.</p> <ul style="list-style-type: none"> • ‘금비녀’, ‘금반지’의 경우, ‘금’이 태깅 대상이 될 수 있는데 이럴 경우 개체 명의 단위가 형태 분석 단위보다 작아 문제가 됨. • ‘산소마스크’의 경우도 ‘산소-마스크’로 <우리말샘> 사전에 등재되어 있고, ‘금비녀’ 또한 <우리말샘> 사전에 ‘금-비녀’로 등재되어 있음. • 따라서, 우리말샘 사전에 ‘^’ 기호를 사용하여 등재된 경우에만 내부 태깅하기로 함. <p>2. AFA_DOCUMENT</p> <ul style="list-style-type: none"> • 정확한 도서명만이 태깅 대상이 됨. <p>3. ‘시스템’의 경우, TR_SCIENCE와 TMI_SW 모두 가능하여 명확한 기준이 필요함.</p> <ul style="list-style-type: none"> • 학문으로도 볼 수 있고, 기술로도 볼 수 있음. • ‘~시스템’은 문맥상 소프트웨어 프로그램이 분명할 때만 TMI_SW로 태깅하고, 어떠한 운용 방법을 의미할 때는 TR_SCIENCE로 태깅함. <p>4. TMM_DISEASE</p> <ul style="list-style-type: none"> • ‘알코올중독, 니코틴중독’ 등은 TMM_DISEASE로 태깅함. ‘스마트폰 중독’의 정의를 보면 ‘~ 병적인 상태’임. 다른 중독과의 일관성을 위해 태깅 대상으로 봄.
3차	국립국어원 데이터리 한림대학교	<p>1. 보통명사의 태깅 단위의 문제</p> <ul style="list-style-type: none"> • ‘어쿠스틱 발라드’, ‘오픈워터 수영’, ‘사촌 동생, 사촌 누나’, 그리고 여러 음식 명칭 등이 전체가 하나를 의미하며, <우리말샘> 사전에 구 구성으로 등재되어 있는 것이 많음. 또한 사전에 등재되어 있지 않더라도 비슷한 구조의 구 구성이 사전에 등재되어 있어 같은 맥락으로 전체를 하나로 태깅함. • 하지만 ‘직업’의 경우, 분리 태깅하고 있어 문제가 됨. • 따라서 구 구성으로 <우리말샘> 사전에 등재되어 있는 경우, 전체를 하나로 보고 CV_OCCUPATION으로 태깅하기로 함.

<표 20> 지침 회의 및 협의 내용

□ 세부 분류별 수정사항

위와 같은 과정을 통하여 최종적으로 각 세부 분류별 분석 지침을 구체적으로 정비하였다.

1) PERSON

- 이전 지침에서 접미사 ‘-이’가 이름에 붙여 쓰인 경우 이를 포함하여

태깅한다고 하였는데, 조사 ‘이랑’의 경우 작업자들에게 전체가 조사인지, 접미사인지 판단이 되지 않아 태깅 범위가 모호하였다. 조사 ‘이랑’의 경우, 사전에 등재되어 있는 조사이므로 태깅 대상이 아님을 추가 기술하였다.

2) STUDY_FIELD

- 수정 내용 없음.

3) THEORY

- TR_OTHERS의 분석 대상인 ‘샤머니즘’ 등은 문화인류학 또는 인류학 전문 용어에 해당하므로 TR_SOCIAL_SCIENCE로 이동하였다. TR_OTHERS는 TR 계열에 속하지 않는 기타 이론/법칙/원리인 ‘교통류 이론, 자동초록기법, 텍스트 마이닝 기법’ 등을 그 분석 대상으로 하였다.

4) ARTIFACTS

- AF_MUSICAL_INSTRUMENT의 분석 대상에 ‘전자악기’를 추가하였다.
- AFA_VIDEO의 주의 사항에 프로그램명과 시리즈나 시즌이 함께 있는 경우는 포함하여 태깅한다는 내용을 추가하였다. 다만 프로그램명이 쓰이지 않고 단독으로 쓰인 ‘시즌, 시리즈’의 경우는 QT_ORDER로 태깅하기로 하였다.

5) ORGANIZATION

- OGG_EDUCATION의 학과나 단과대학은 학교에 포함되어 있기 때문에 OGG_EDUCATION으로 태깅하지 않는다는 주의 사항의 내용에 ‘한양공대-<한양: OGG_EDUCATION> 공대’ 예를 추가하였다. 2020년 지침에는 ‘홍익대 미대’라는 예문이 있는데, 붙여서 쓰일 수 있는 예문을 제시하여 자세히 설명하였다.

6) LOCATION

- LC_OTHERS의 분석 대상에 ‘신도시, 기업도시’ 등 도시의 특정 구역

을 나타내는 명칭을 추가하였다.

- LCG_CONTINENT의 정의 및 설명에 ‘대륙의 일부분의 명칭’을 추가하였다. 또한 분석 대상에 ‘중동’을 추가하였다.

7) CIVILIZATION

- ‘기관+직위’의 경우 분리가 되어 태깅이 가능한 경우, ‘기관’은 OG G 계열로, ‘직위’는 CV_POSITION으로 태깅하기로 하였다. ‘남구청 부구청장’은 분리가 되어 태깅이 가능하며, <우리말샘> 사전에도 구 구성으로 등재되어 있지 않다. 따라서 ‘남구청 부구청장’은 분리하여 태깅한다. 다만 ‘선거대책위원장’과 같이 기관과 직위가 분리되지 않는 경우는 전체를 CV_POSITION으로 태깅하는 것은 유지하였다.
- ‘직업’의 경우, 예를 들어 ‘힙합 댄서’, ‘상송 가수’는 <우리말샘> 사전에 ‘힙합 댄서’, ‘상송 가수’와 같이 ‘구’로 등재되어 있다. 또한 사전에 등재되어 있지 않더라도 구 구성의 내용이 직업의 명칭으로 인정된다면 전체가 하나의 직업 명칭을 의미한다. 따라서 ‘최장 단위 태깅’을 하여 전체를 하나로 태깅한다. 2020년 작업에서는 ‘최장 단위 결정의 원리-보통명사’에 ‘사촌 동생’과 같은 구 구성의 경우는 최장 단위 태깅을 하였는데, ‘직업’의 경우에는 구 구성을 최소 단위 태깅을 하였다. 2021년 작업에선 ‘직업’, ‘직위’, ‘가족’의 경우 <우리말샘> 사전의 등재 방식을 기준으로 삼아 지침에 일관성을 확보하였다.

8) DATE

- DT_MONTH의 분석 대상에 ‘가정의 달’, ‘호국보훈의 달’ 등 특정 달을 지칭하는 명칭을 추가하였다.
- 각각의 소분류 주의 사항에 ‘약, 여, 쯔’ 등의 함께 태깅할 수 있는 구체적 표현을 제시하였다.

9) TIME

- 각각의 소분류 주의 사항에 ‘약, 여, 쯔’ 등의 함께 태깅할 수 있는

구체적 표현을 제시하였다.

10) QUANTITY

- 수학적 기호가 포함되어 있을 때 기존의 지침에는 ‘숫자+수학적 기호+숫자’의 경우는 태깅 대상이 되나 특정 수치를 가리킬 때, 특정 태그에 태깅하거나 OTHERS에 태깅하는 등 그 기준이 모호했다. 따라서 다음과 같이 그 기준을 명확히 하였다. 첫째, ‘수학적 기호+숫자’의 경우 태깅하지 않는다. 둘째, ‘숫자+수학적 기호+숫자’의 경우는 결핍값이 확정되므로 전체를 QT_OTHERS로 태깅하기로 하였다.

11) EVENT

- EV_FESTIVAL의 분석 대상에 ‘박람회/엑스포’ 명칭을 추가하였다.

12) ANIMAL

- 수정 내용 없음.

13) PLANT

- 수정 내용 없음.

14) MATERIAL

- MT 계열은 해당 원소명 및 금속물 등 물질의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다. <우리말샘> 사전에 ‘^’로 연결되어 있는 경우, MT에 한해 ‘^’를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)하기로 하였다.

15) TERM

- ‘알코올 중독, 니코틴 중독’ 등은 TMM_DISEASE로 태깅하였다. ‘스마트폰 중독’의 정의를 보면 ‘~ 병적인 상태’로 다른 중독과의 일관성을 위해 ‘스마트폰 중독’을 분석 대상에 추가하였다.

2. 개체 연결 말뭉치 구축 지침 수립

2.1. 개체 연결 말뭉치 구축 지침 수립 과정

‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’에서는 개체 연결 말뭉치 구축을 위한 지침 수립이 주요 과제 중 하나였다. 개체 연결 말뭉치 구축 지침을 수립하기 위해 위키피디아의 체계, 등재 기준, 등재 형식 등을 분석함과 동시에 실제 개체 연결을 진행하며 발견되는 사례들을 유형화, 체계화하여 지침에 반영될 수 있도록 하였다.

개체 연결 작업 전에는 이번 사업에서 개체 연결을 진행할 대분류인 PERSON, ARTIFACTS, ORGANIZATION, LOCATION, DATE에 해당하는 개체명들을 기존 2020년 개체명 분석 말뭉치 구축 결과물을 토대로 추출하고, 위키피디아에의 등재 여부, 등재 방식 등을 분석하였다. 이러한 과정을 통해 개체명과 위키피디아 문서를 연결할 때에 발생할 수 있는 문제점들을 정리하였고, 이에 대한 연결 방법을 수립하여 개체 연결 말뭉치 구축 지침 초안을 작성하였다.

개체 연결 작업 중에는 실제로 개체 연결을 하면서 나타나는 난점, 의문점 등이 공유될 수 있도록 작업자들에게 구글 시트에 이를 기재하도록 하였다. 그중 주요 사항들은 회의를 통해 연결 방안을 논의 후, 결정된 사항을 지침에 반영하였다. 이 작업은 개체 연결을 진행하는 내내 이루어졌으며, 이 과정을 통해 지속적으로 지침의 수정도 이루어졌다.

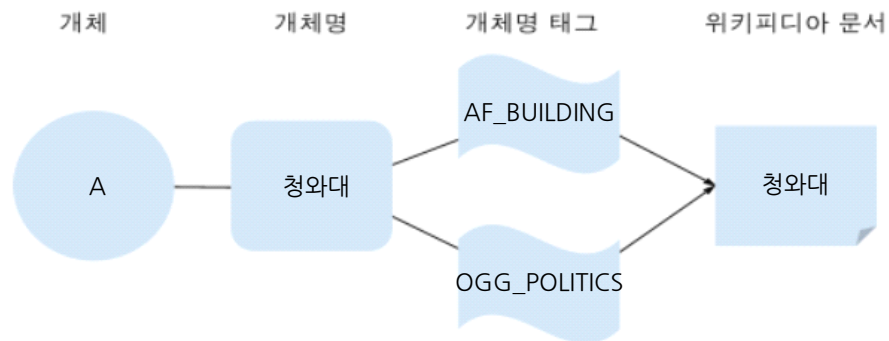
개체 연결 말뭉치 구축 지침과 지침을 토대로 한 개체 연결의 결과물을 국립국어원, 데이터리, 한림대학교가 여러 차례 검토하였으며, 이 과정에서 드러나는 문제점을 세 기관이 논의한 후 협의 결과를 지침에 반영하였다. 지침이 수정된 경우 개체 연결의 수정도 이루어졌다.

<2021년 개체 연결 말뭉치 구축 지침>은 위와 같은 과정을 통해 1.1. 버전으로 시작하여 최종 1.6. 버전으로 마무리되었다.

2.2. 개체명과 지식베이스(위키피디아) 문서 간의 관계 유형

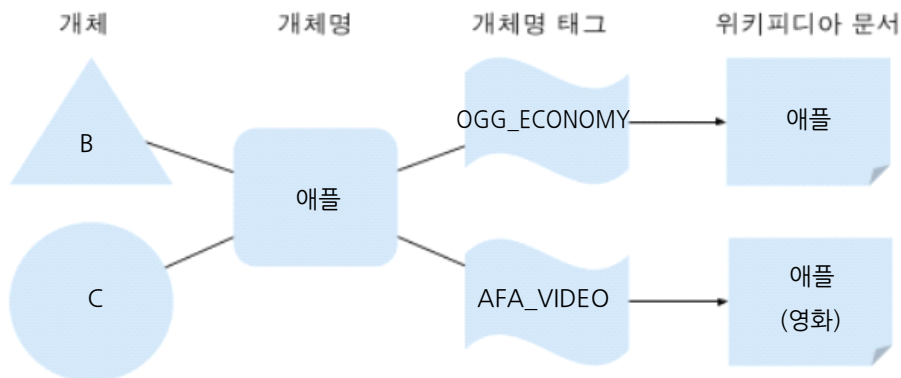
<2021년 개체 연결 말뭉치 구축 지침>에서는 개체명과 위키피디아 문서를 올바르게 연결하기 위하여 개체(실제 세계의 대상), 개체명(언어 표현)과 개체명 태그(범주적 성격 표지), 그리고 위키피디아 문서(백과사전적 정보) 간의 관계를 개체명의 어휘 관계를 기준으로 1. 다의어, 2. 동음이의어, 3. 상위어-하위어, 4. 전체어-부분어, 5. 동일지시적 표현 등 다섯 가지로 유형화하였으며, 이를 도식으로 표현하고 각각의 사례를 제시하였다. 각 유형별 도식과 그에 해당하는 사례는 아래와 같다.

1) 다의어

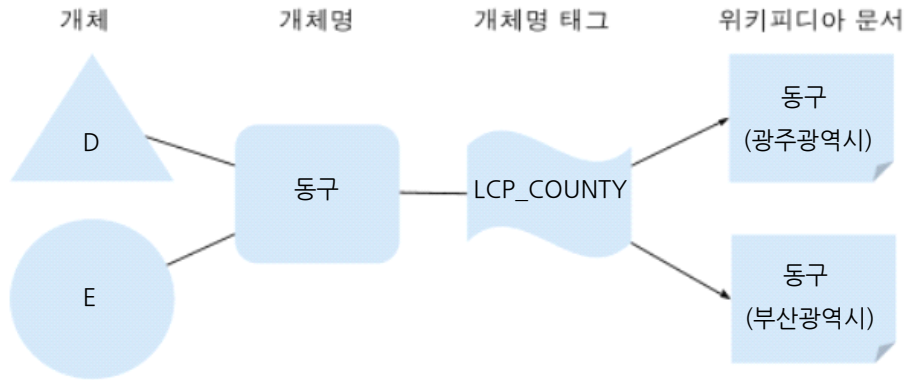


<그림 17> 다의어 도식화

2) 동음이의어

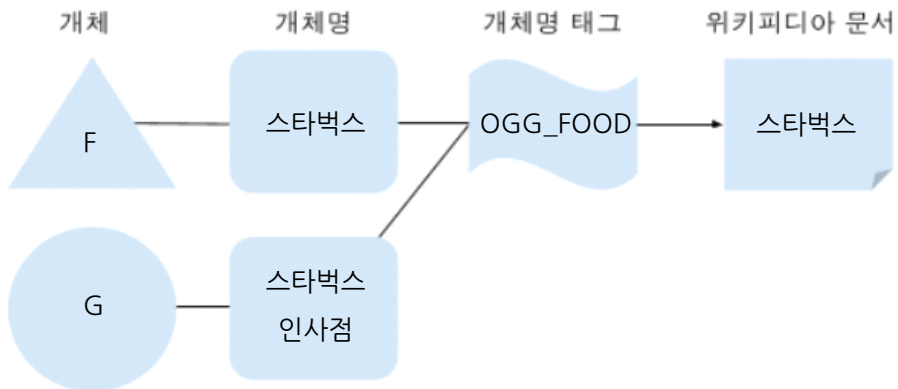


<그림 18> 동음이의어 도식화 1



<그림 19> 동음이의어 도식화 2 (개체명 태그가 같은 경우)

3) 상위어-하위어

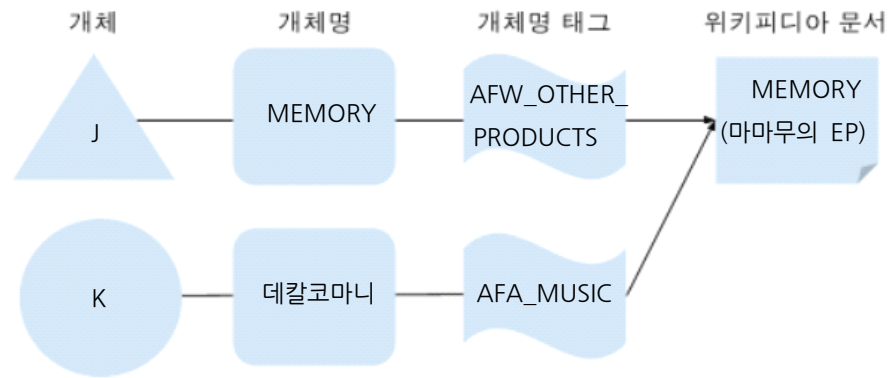


<그림 20> 상위어-하위어 도식화 1 (하위어 문서가 별도로 존재하지 않는 경우)

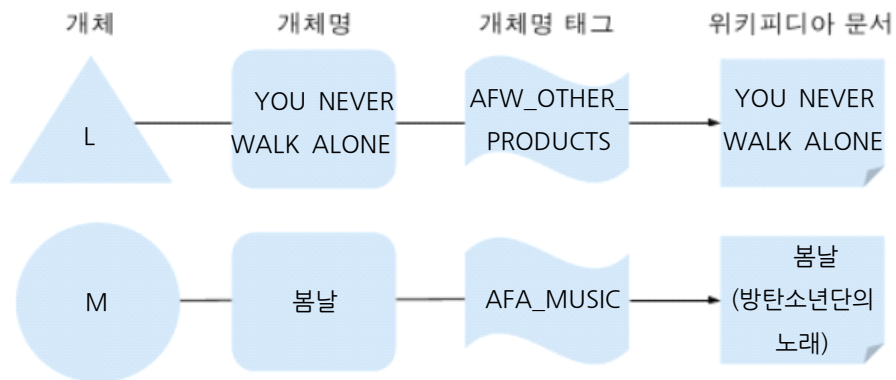


<그림 21> 상위어-하위어 도식화 2 (하위어 문서가 별도로 존재하는 경우)

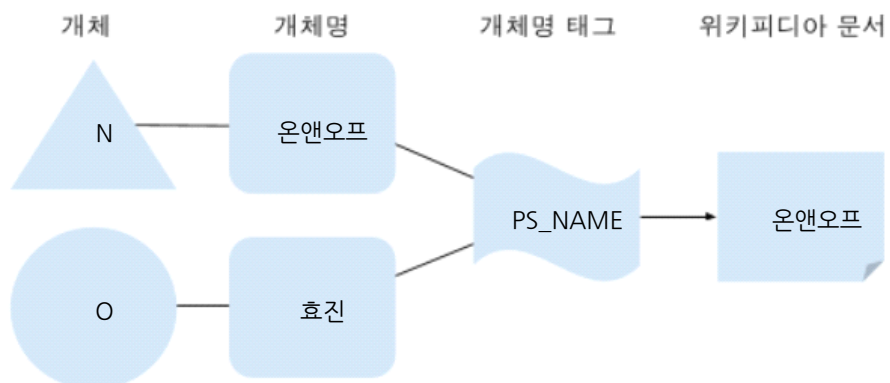
4) 전체어-부분어



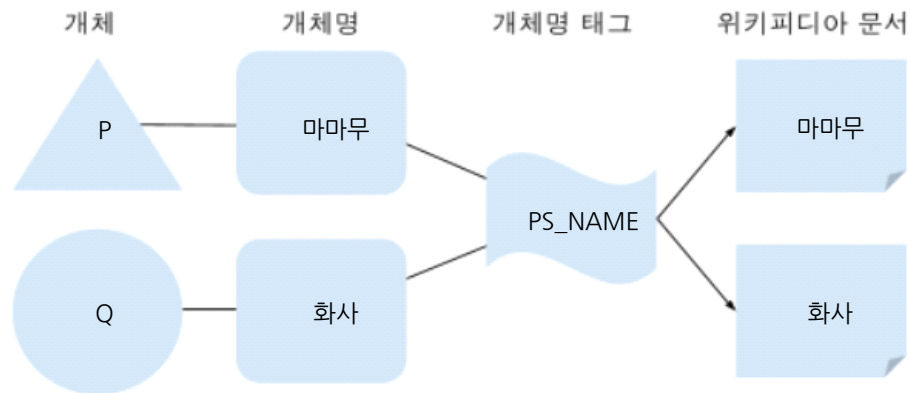
<그림 22> 전체어-부분어 도식화 1
(개체명 태그가 다르며 전체어 문서에서 부분어 정보를 다루고 있는 경우)



<그림 23> 전체어-부분어 도식화 2
(개체명 태그가 다르며 전체어 문서에서 부분어 문서가 각각 존재하는 경우)

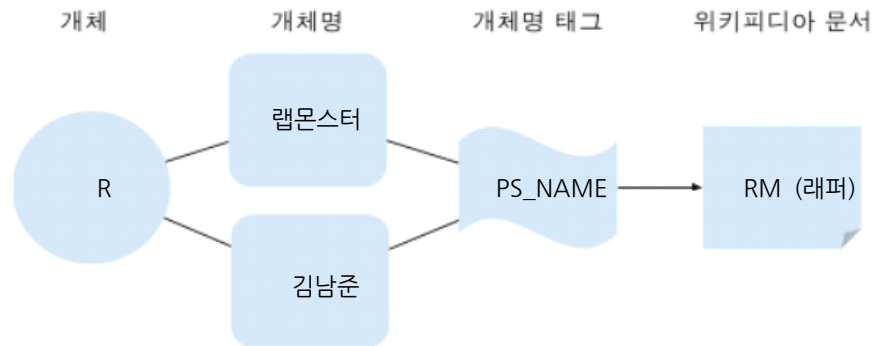


<그림 24> 전체어-부분어 도식화 3
(개체명 태그가 같으며 전체어 문서에서 부분어 정보를 다루고 있는 경우)



<그림 25> 전체어-부분어 도식화 4
(개체명 태그가 같으며 전체어 문서와 부분어 문서가 각각 존재하는 경우)

5) 동일지시적 표현



<그림 26> 동일지시적 표현

2.3. 개체명과 지식베이스(위키피디아) 문서 연결의 원칙

2.3.1. 대원칙

개체명과 위키피디아 문서의 연결은 연결 대상, 연결 단위, 비연결 항목의 세 부분으로 나누어 대원칙을 마련하였다. 대원칙은 세부 원칙에도 일관적으로 적용되는 것으로, 개체명과 위키피디아 문서 연결의 대원칙의 주요 내용은 아래와 같다.

첫째, 연결 대상은 개체명의 언어 표현과 동일한 제목의 ‘한국어판 위키피디아 백과사전’의 문서이며, ‘한국어판 위키피디아 백과사전’에 개체명과 일치하는 제목의 문서가 없는 경우, ‘영어판 위키피디아 백과사전’의 문서를 연결

하도록 하였다. 문서 연결의 일차적 기준은 ‘제목’ 이나, 개체명이 동일지시적 표현을 다수 가질 수 있으므로 언어 표현이 다르더라도 동일한 대상(개체)에 대한 문서이면 연결하도록 하였다. 또한 개체명 태그와 동일한 범주에 속하는 위키피디아 문서를 연결하도록 하여 동음이의 관계에 있는 개체명들이 올바르게 연결될 수 있도록 하였다.

둘째, 연결 단위는 국립국어원의 <2020년 개체명 분석 말뭉치 구축 지침>의 개체명 태깅 단위를 따르는 것을 원칙으로 하였으며 개체명 설정 단위 전체와 일치하는 제목의 위키피디아 문서가 존재하지 않는 경우 문서를 연결하지 않는 것으로 하였다. 다만, ‘스타벅스 인사점’ 과 같이 상위 조직과 하위 지점(지부)으로 구성된 개체명의 경우에 한해 예외적으로 구성 전체(스타벅스 인사점)가 별도의 위키피디아 문서로 존재하지 않는다면 상위 조직 문서(스타벅스)를 연결하도록 하였다.

셋째, 비연결 항목은 개체명에 위키피디아 문서를 연결하지 않는 항목으로 아래에 제시한 목록이 이에 해당한다.

- (1) 개체명과 동일한 제목의 위키피디아 문서가 존재하지 않는 경우
- (2) 개체명과 동일한 제목의 위키피디아 문서가 존재하나 의미가 다른 경우
- (3) 개체명이 동음이의어인데 대상 문서에서 개체명이 지시하는 대상을 맥락을 통해 특정할 수 없는 경우
- (4) 개체명이 하위어인데, 위키피디아에 하위어 문서가 별도로 존재하지 않고 상위어 문서로 넘어가는 경우
- (5) 개체명이 부분어인데, 위키피디아에 부분어 문서가 별도로 존재하지 않고 전체어 문서로 넘어가는 경우
- (6) 개체명 설정 단위 전체와 일치하는 제목의 위키피디아 문서가 존재하지 않는 경우
- (7) 개체명의 본유적 의미와 맞지 않는 개체명 태그가 부착되어 있는 경우

2.3.2. 개체명의 어휘 관계에 따른 세부 원칙

개체명과 위키피디아 문서 연결의 세부 원칙을 개체명의 어휘 관계에 따라 1. 개체명이 다의어인 경우, 2. 개체명이 동음이의어를 가지는 경우, 3. 개체명이 상위어를 가지는 경우, 4. 개체명이 전체어를 가지는 경우, 5. 개체명이 동일지시적 표현을 가지는 경우 등 다섯 가지로 나누어 마련하였다.

1) 개체명이 다의어인 경우

- 개체명이 다의어인데, 하나의 위키피디아 문서 안에 둘 이상의 의미가 모두 기술되어 있는 경우 서로 다른 세분류로 태깅된 개체명에 해당 위키피디아 문서를 각각 연결하도록 하였다.
- 개체명이 다의어인데, 하나의 위키피디아 문서 안에 하나의 의미만 기술되어 있는 경우에도 해당 위키피디아 문서를 서로 다른 세분류로 태깅된 개체명에 각각 연결하도록 하였다.

2) 개체명이 동음이의어를 가지는 경우

- 개체명이 동음이의어를 가지는 경우, 즉 개체명이 지시하는 대상이 여럿 존재하는 경우에는 개체명과 동일한 제목을 지닌 위키피디아 문서 중에서 괄호 안의 추가 정보와 본문의 내용으로 보았을 때 개체명이 지시하는 대상과 일치하는 문서를 찾아 연결하도록 하였다.

3) 개체명이 상위어를 가지는 경우

- 개체명이 하위어인데, 위키피디아에서 해당 개체명을 검색하였을 때 상위어 문서로 자동으로 넘어가는 경우, 연결하지 않는 것을 원칙으로 하였다.
- 다수의 지점 혹은 지부를 보유하고 있는 기업이나 기관의 경우, 개체명이 하위 조직을 의미하는데 해당 하위 조직이 별도의 문서로 존재하지 않을 때는 예외적으로 상위 조직 문서를 연결하도록 하였다.

4) 개체명이 전체어를 가지는 경우

- 개체명이 부분어인데, 위키피디아에서 해당 개체명을 검색하였을 때 전

체어 문서로 자동으로 넘어가는 경우, 연결하지 않는 것을 원칙으로 하였다.

- 시즌제 TV 프로그램이나 드라마의 경우에는 각각의 시즌이 별도의 문서로 존재하지 않고 대표 문서 안에서 각각의 시즌을 기술하고 있다면, 예외적으로 해당 시즌을 대표 문서에 연결하도록 하였다.

5) 개체명이 동일지시적 표현을 가지는 경우

- 정식 명칭의 일부만 표현된 개체명, 줄임말, 일상적으로 부르는 말 등 하나의 대상을 지칭하는 표현이 다수 존재하는 경우에는 해당 대상에 대해 위키피디아에서 제공하는 정식 명칭의 문서를 연결하도록 하였다.

2.3.3. 개체명의 대분류에 따른 세부 원칙

개체명과 위키피디아 문서 연결의 세부 원칙을 개체명의 대분류에 따라 1) PERSON, 2) ARTIFACTS, 3) ORGANIZATION, 4) LOCATION, 5) DATE 등 다섯 가지로 나누어 마련하였다.

1) PERSON

- 개체명이 ‘성 + 직위/직책’, ‘성 + 씨’ 구성인 경우, 맥락을 통해 개체명이 지시하는 대상을 명확하게 판별할 수 있다면 개체명 ‘성’에 해당 성명(성+이름)을 제목으로 하는 위키피디아 문서를 연결하도록 하였다.
- SNS 계정명의 경우, 실제 인물에 대한 위키피디아 문서와 SNS 계정명을 연결하지 않는 것으로 하였다.
- 어떤 인물이 음악 그룹에 속해 있는데, 위키피디아에 인물에 대한 문서가 별도로 존재하지 않고 음악 그룹 문서로 넘어가는 경우에는 개체명 ‘인물’에 위키피디아 ‘음악 그룹’ 문서를 연결하지 않는 것으로 하였다.
- PS_CHARACTER로 태깅된 실존 인물의 경우, 작품 속 인물에 대한 문서가 위키피디아에 별도로 존재하지 않을 때 실존 인물 문서를 연결하지

않는 것으로 하였다.

2) ARTIFACTS

- 동명의 소설, 그림, 영화, 음악 등이 존재하는 경우, 즉 서로 다른 예술 분야에 동명의 작품이 존재하는 경우에는 개체명 태깅 범주와 동일한 예술 분야의 문서를 연결하도록 하였다.
- 영화 ‘스타워즈’, ‘어벤져스’ 등과 같이 시리즈로 제작된 영화의 경우, 영화의 특정 시리즈가 위키피디아에 별도의 문서로 존재하는 경우에는 해당 개체명과 위키피디아 문서를 연결하고 특정 시리즈가 위키피디아에 별도의 문서로 존재하지 않는 경우에는 연결하지 않는 것으로 하였다.
- 미국 드라마 ‘프렌즈’ 나 예능 프로그램 ‘신서유기’, ‘1박 2일’ 등과 같이 시즌제로 제작되는 TV 프로그램이나 드라마의 경우, 각각의 시즌이 별도의 문서로 존재하지 않고 하나의 문서 안에 관련 정보로서 포함되어 있는 경우, 예외적으로 해당 시리즈를 대표 문서에 연결하도록 하였다.
- 노래 제목을 위키피디아에 검색하였을 때, 해당 노래를 주제로 하는 위키피디아 문서가 존재하지 않고 해당 노래가 수록되어 있는 음반으로 문서가 넘어가는 경우, 곡명은 AFA_MUSIC, 음반 명은 AFW_OTHER_PRODUCTS로 개체명 태그가 다르므로 개체명 ‘노래 제목’에 위키피디아의 ‘음반’ 문서를 연결하지 않는 것으로 하였다.
- 기업의 하위 지점이나 지부가 기관(OGG)의 의미를 가지는데, 해당 하위 지점/지부가 별도의 문서로 존재하지 않고 상위 기업 문서만 존재하는 경우에는 상위 기업 문서를 연결하나, 건물(AF)의 의미를 가질 때에는 상위 기업 문서를 연결하지 않는 것으로 하였다.

3) ORGANIZATION

- 특정 기업이 보유하고 있는 브랜드가 위키피디아에 별도의 문서로 존재하지 않고 상위 기업 문서로 넘어가는 경우, 개체명 ‘하위 브랜드’에 위키피디아 ‘상위 기업’ 문서를 연결하지 않는 것으로 하였다.

- 기관이나 기업이 통폐합 혹은 개편되어 명칭이 변경된 경우, 변경되기 이전의 개체명에 변경된 이후의 명칭을 제목으로 하는 위키피디아 문서를 연결하지 않는 것으로 하였다.
- 국가명이 LCP_COUNTRY가 아니라 OGG_SPORTS로 태깅되어 있는 경우에는 위키피디아의 ‘국가’ 문서가 아닌 ‘국가대표팀’ 문서를 연결하도록 하였다.

4) LOCATION

- 대상 문서에서 태깅된 개체명의 태그와 동일한 주제의 위키피디아 문서를 연결하도록 하였다. 즉, 문서에서 제주도가 LCP_PROVINCE로 태깅되어 있으면 행정구역인 제주도를 주제로 하는 위키피디아 문서, LCG_ISLAND로 태깅되어 있으면 섬인 제주도를 주제로 하는 위키피디아 문서를 연결하도록 하였다.
- 상위의 행정구역명과 하위의 행정구역명이 동일한데 개체명 태그도 동일한 경우, 대상 문서의 맥락을 통해 어떤 행정구역인지 판별하고, 그에 맞는 위키피디아 문서를 연결하도록 하였다. 맥락을 통해 대상 개체명의 행정구역을 명확하게 판별할 수 없는 경우, 상위의 행정구역에 대한 위키피디아 문서를 우선적으로 연결하도록 하였다.
- 위키피디아에서 법정동을 검색하였을 때 법정동 문서는 존재하지 않고 행정동 문서로 넘어가는 경우, 개체명 ‘법정동’에 위키피디아 ‘행정동’ 문서를 연결하지 않도록 하였다.
- 과거의 지명을 위키피디아에서 검색하였을 때, 과거 지명 문서는 존재하지 않고 현재 지명 문서로 넘어가는 경우 연결하지 않는 것으로 하였다.

5) DATE

- ‘초여름’, ‘수요일부터’, ‘2002년경’, ‘이번 주말’ 등과 같이 날짜나 계절 표현에 접사, 조사, 관형사 등이 결합되어 있으나 해당 언어 표현과 동일한 위키피디아 문서가 존재하지 않는 경우, 위키피디아 문서를 연결하지 않는 것으로 하였다.
- ‘국가명 + 초기, 중기, 말기’와 같은 시기 구분 명칭의 경우, 해당 구

성 전체가 위키피디아에 별도의 문서로 존재하지 않고 ‘국가’ 문서로 자동으로 넘어가더라도, 개체명 ‘국가명 + 초기, 중기, 말기’에 위키 피디아 ‘국가’ 문서를 연결하지 않는 것으로 하였다.

2.4. 개체 연결 말뭉치 구축 지침 정비

2.4.1. 자문 내용

개체 연결 말뭉치 구축 지침의 완성도를 높이기 위하여 국어학 전문가와 빅 데이터 전문가에게 자문하였다. 자문 내용은 아래와 같다.

자문 위원	아주대학교 박재연 교수님
자문 내용	<ol style="list-style-type: none"> 1. 개체명이 하위어인데, 위키피디아 검색 시 상위어 문서로 넘어가는 경우 → 이러한 예들을 다 연결하게 되면 그 한계를 어디까지 정해야 하는가에 대한 어려움이 발생할 수 있다. 2. 개체명이 부분어인데, 위키피디아 검색 시 전체어 문서로 넘어가는 경우 → 시즌제 드라마, 영화 시리즈물과 같은 부류는 연결을 고민해 볼 필요가 있다. 3. '성 + 직위/직책' 구성인 개체명 → 정확한 정보를 얻을 수 없다는 문제가 발생하기는 하나, 연결할 수밖에 없다. 4. AF로 태깅된 하위 지점/지부가 있는 기업 → 현 지침의 처리 방안에 대해 이견이 없다. 5. 통폐합·개편되어 명칭이 변경된 기관이나 기업 → 정보를 얻을 수 있다는 관점에서 인간에게는 매우 유의미한 정보이지만 어디까지 연결할 것인지 한계를 정하기가 어렵다. 이러한 한계에 대해 언어학적으로 경계를 주기에는 어려운 문제라고 판단되기에, 실용성과 실현 가능성의 관점에서 접근해야 한다고 생각한다. 6. 개체명이 법정동인데, 위키피디아 검색 시 행정동 문서로 넘어가는 경우 → 지시 대상은 다르나, 실용적인 목적에서는 중요하기에 연결해야 한다고 생각한다.

<표 21> 전문가 자문 내용 1

자문 위원	숙명여자대학교 이홍식 교수님
자문 내용	<p>1. 개체명이 하위어인데, 위키피디아 검색 시 상위어 문서로 넘어가는 경우 → 목적에 따라 연결 여부가 달라질 수 있다. 인간이 검색을 통해 무엇을 얻고자 한다면 이와 같은 연결이 도움이 될 수 있으나, 기계적 입장에서 맞지 않는 내용이 포함되기 때문에 처리 과정에 문제가 생길 수 있다. 정확한 정보만을 연결하는 것이 후처리 과정에서 더욱 효율적이라고 생각된다.</p> <p>2. 개체명이 부분어인데, 위키피디아 검색 시 전체어 문서로 넘어가는 경우 → 기계적인 입장에서 이러한 연결이 불필요하다고 본다. 올해 사업에서는 최소주의 원칙으로 정교하게 작업하고 추후에 어떤 부분에서 정보가 부족한지 파악한 후 보완하는 것이 나을 것이다.</p> <p>3. '성 + 직위/직책' 구성인 개체명 → 현 지침과 같이 개체명 태깅 단위를 기준으로 연결하는 것이 옳다고 본다. 필요하다면 별도의 목록으로 작성할 수 있다.</p> <p>4. AF로 태깅된 하위 지점/지부가 있는 기업 → 사용자가 '기업의 특정 하위 지점/지부(AF)'를 검색했을 때, '본사(OG)'가 나온다면 사용자에게는 불필요한 정보이다. 따라서 일관성을 유지하기 위해 '건물/빌딩'의 의미일 때는 연결하지 않는 것이 좋을 듯하다.</p> <p>5. 통폐합·개편되어 명칭이 변경된 기관이나 기업 → 최소주의 입장에서 연결하지 않는 것이 타당하다. 추후에 필요하다면 추가로 연결할 수 있다.</p> <p>6. 개체명이 법정동인데, 위키피디아 검색 시 행정동 문서로 넘어가는 경우 → 법정동과 행정동의 지시 대상이 완전히 다르다기보다는 포함되거나 비슷한 경우가 있고, 연결을 안 하게 되면 정보를 전혀 가져올 수 없기에 연결해야 한다고 생각한다.</p>

<표 22> 전문가 자문 내용 2

자문 위원	카이스트 최기선 교수님
자문 내용	<p>1. 개체명-위키피디아 문서 연결 말뭉치가 활용될 수 있는 응용 분야 및 활용의 측면에서 개체명 연결 시 고려해야 할 점 → 1) 위키피디아는 일반 커뮤니티에서 등재어를 올리고, 등재어는 고유명사가 대부분이다. (물론 일반명사도 들어있지만 특별한 의미를 설명할 필요가 있을 경우 올라간다고 보인다.) 개체명이 고유명사를 대상으로 하고, 위키피디아의 등재어를 대상으로 하면 자연적으로 개체명의 시작-끝의 범위에 대한 일반 보편적 데이터를 얻을 수 있다.</p>

2) 정제된 데이터로서의 위키피디아 데이터 사용이 가능하다. 개체명에 대한 문맥을 한정적인 데이터 내에서 다양하게 모을 수 있다.

3) 위키피디아 문서 중 사건 이름 (예; 대한민국의_대통령_선거)인 경우, 개체명 최대범위와 부분 범위에 대한 기준이 필요하다.

2. <2021년 개체 연결 말뭉치 구축 지침>에서는 개체명이 지시하는 대상과 위키피디아 문서에서 설명하는 대상이 일치하는 때에만 이 둘을 연결하는 것을 대원칙으로 삼고 있음. 따라서 개체명과 위키피디아 문서가 하위어-상위어 관계(파리바게뜨-파리크라상, 나룻배-선박)일 때, 그리고 부분어-전체어 관계(화사-마마무, 데칼코마니-메모리)일 때에는 이 둘을 연결하지 않고 있음. 이에 대한 의견이 어떠한지?

→ 위키피디아 문서는 URI로서 유일한 개념을 가진 개체이므로, 상·하위, 부분 전체어로서 연결된 개체도 연결할 경우, 유일한 개념이라는 기본적 정의를 위배한다.

3. 개체명 분석 단계에서 두 단위로 나뉘어 태깅되지만 실제로는 하나의 대상을 지시하는 개체명과 위키피디아 문서의 연결에 대한 질문. (문 대통령, 이명박 정부 등)

→ 개체명 인식 말뭉치이고 위키피디아 문서 개체를 기준으로 한다면, 위키피디아 문서 개체에 충실해야 한다. 위키피디아 문서는 일반 커뮤니티가 원하고 보편적이므로 이에 위반하는 것은 커뮤니티의 흐름에 어긋난다. 개체명 인식의 보편적 접근이 일반인이라고 한다면 더욱 그러하다. 복합명사 내부의 단일명사는 각각에 대해서는 그 복합명사 전체 의미를 상실하였다고 보는 것이 맞다. 엄중하게 태깅한다면 전체 복합 개체명도 태깅을 하고 부분 개체명이 위키피디아에 있다면 그것도 하도록 하는 것이 좋겠다. 테스트할 경우, 전체 점수와 부분 점수를 따로 측정하는 방안이 이미 논문에 나와 있다.

4. 엑소브레인에서 지식베이스 기반 추론 기술을 총괄하며 경험한 데이터와 본 사업의 개체 연결 방식 비교하였을 때 어떤 차이점이 있는지?

→ 엑소브레인에서는 URI로서의 개체 연결에 충실하였으므로, Wikipedia의 데이터 웹 버전인 DBpedia의 모든 개체에 접근할 수 있느냐가 문제가 된다. 따라서 위 3번과 같은 경우, 찾을 수 없게 된다. 따라서, 개체명 인식 결과를 보고, 개체 연결 연구를 더 해야 하는 이유가 되었다.

5. 엑소브레인 지식 연결을 연구를 진행하며 가장 어려운 점이 무엇이었는가? 또는 차후 연구에 계획한 항목 중 저희 위키피디아 데이터 구축과 연관된 것이 있는지?

→ 첫째, 개체 연결을 꼭 해야 하는 방향으로 하다 보니, 위키피디아에 없는 개체인지 아닌지를 판별하는 NA 문제가 효율성을 낮추었다. 이에 대한 데이터 대책이 필요하다.

둘째, 사건명 혹은 명사구 개체명 (예: 대한민국의 대통령 선거, 조선 인민민주공화국의 대량 살상무기)를 찾지 못하는 문제 혹은 텍스트 내의 개체명 범위 문제를 풀 수 있는 데이터가 필요하다.

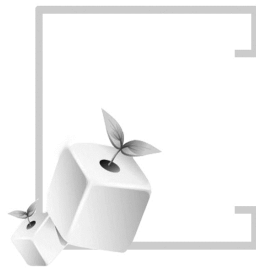
<표 23> 전문가 자문 내용 3

2.4.2. 협의 내용

지침 수정을 위해 기관 간 협의가 이루어졌다. 세부 내용은 아래와 같다.

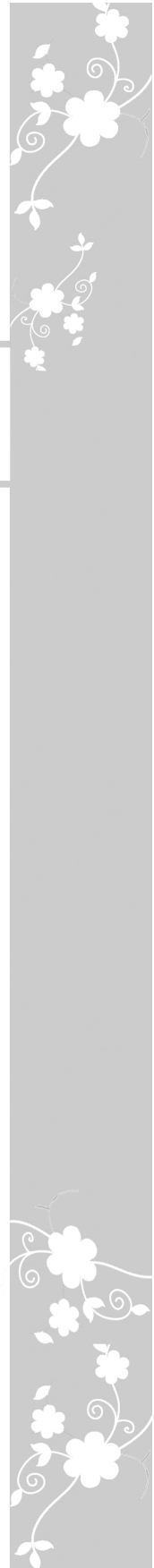
차수	참여 기관	협의 내용
1차	국립국어원 데이터리 한림대학교	<ol style="list-style-type: none"> PERSON <ul style="list-style-type: none"> (개체명) 수호랑/PS_CHARACTER - (위키피디아) 수호랑과 반다비 연결 여부 ORGANIZATION <ul style="list-style-type: none"> 〈대한민국: LCP_COUNTRY〉 〈정부: OGG_POLITICS〉와 같이 지시 대상은 하나이나, 분리 태깅된 개체명의 연결 문제 LOCATION <ul style="list-style-type: none"> (개체명) 지곡동(법정동)/LCP_COUNTY - (위키피디아) 보라동 (행정동) 연결 여부 DATE <ul style="list-style-type: none"> (개체명) 고려 말기/DT_DYNASTY - (위키피디아) 고려 연결 여부 (개체명) 1월 말/DT_MONTH - (위키피디아) 1월 연결 여부
2차	국립국어원 데이터리 한림대학교	<ol style="list-style-type: none"> 개체명이 비유적 의미를 지니는 경우에 대한 연결 문제 <ul style="list-style-type: none"> 지난주에 〈아프리카: LCG_CONTINENT〉 갔다 왔다. 대구 진짜 덥더라. 대프리카 인정. (개체명) 아프리카: LCG_CONTINENT - (위키피디아) 아프리카 개체명의 본유적 의미와 맞지 않는 개체명 태그가 부착되어 있는 경우에 대한 연결 문제 <ul style="list-style-type: none"> 내가 상당히 어려워하는 작곡가 〈모차르트: PS_NAME〉 〈피아노 소나타 16번: AFA_MUSIC〉. 정말 〈모차르트: AFA_MUSIC〉 너무 어렵다.

<표 24> 지침 회의 및 협의 내용 2



제 4 장

결론



1. 사업 요약

국립국어원이 주관하고 기업 데이터리(Data.ly)와 공동 수행 기관 한림대학교가 수행한 ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업은 약 7개월간 진행되었다. 사업의 주 내용은 문어/구어 원시 말뭉치 300만 어절과 ‘국립국어원 개체명 분석 말뭉치 (버전 1.0)’ 300만 어절에 개체명 세분류 150개를 기준으로 표지를 부착하는 작업을 진행하여 성능 97%에 도달하는 말뭉치를 구축하는 것과 작년 개체명 사업을 통해 완성된 ‘20년 국립국어원 개체명 분석 말뭉치 지침’을 재정비하여 지침의 명확성과 완성도를 높이는 것이었다. 또한, 개체 연결이라는 신규 과업이 추가되어 개체명 분석 말뭉치를 대상으로 문서 내 PERSON, ARTIFACTS, ORGANIZATION, LOCATION, DATE 표지가 부착되어 있는 개체 표현이 의미적으로 지칭하는 개체를 위키피디아 상에서 찾아 연결 정보를 부착하는 작업을 진행하였고, 이를 위한 ‘개체 연결 분석 지침 1.6 버전’을 기술하여 작업의 기준을 수립하였다. 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 ‘기술팀’과 ‘언어팀’으로 분류하였다. 기술팀에서는 구축 프로그램 ‘케인’, 클라우드 소싱 플랫폼 ‘와이즈돔’을 이용한 개체명 부착 작업과 자동화 시스템을 이용한 오류 검출 및 수정 작업과 인력 관리를 맡았다. 언어팀에서는 국어학 전문가가 지침에 대해 검토하고 문제점을 보완하는 작업과 개체 연결 작업 및 지침 작성 작업을 시행하였다.

말뭉치 구축 작업 과정에서는 자체 개발한 개체명 구축 프로세스와 구축 도구 ‘케인’, 클라우드 소싱 플랫폼 ‘와이즈돔’을 병행하여 이용했기 때문에 많은 인원이 투입되었음에도 일정의 지연 없이 효율적으로 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.

지침 수정 및 보완의 과정에서는 20년도 지침 수정 작업에서 마무리 짓지 못했던 세분류 간 내용이 중첩되는 부분과 전반적인 지침의 일관성 문제 등을 해

결하여 명확성을 높이고 150개 세분류의 내용을 상세화하였다. 또한 개체 연결 연구 분석을 통해 구축 작업을 위한 지침 초안을 작성한 뒤 구축 작업을 진행하였으며 작업 중 발생한 문제점 및 추가 내용을 지침에 반영하여 이론에 그치지 않고 실제 구축 작업에 적용할 수 있는 지침을 구축하였다.

2. 의의 및 기대 효과

4차 산업혁명과 비대면 업무의 증가로 여러 분야에서 인공지능 기술 개발에 대한 관심이 집중되고 있으며, 이에 따라 인공지능 기술 개발을 위한 학습데이터에 대한 수요도 점차 증가하고 있다. 하지만 수요에 비해 한국어 데이터의 공급은 충분치 못한 실정이다. 이와 같은 상황에서 본 사업의 결과물인 대용량의 개체명 분석 말뭉치와 개체 연결 분석 말뭉치는 인공지능 기술 개발과 국어 및 국어문화 연구와 국어정책 수립의 기초 자료로 쓰이는 등 매우 중요한 학습데이터로 여러 방면에서 유용하게 활용될 것이다. 또한 재정비되어 완성도를 높인 개체명 분석 지침과 새로 기술된 개체 연결 분석 지침은 말뭉치에 대한 신뢰성을 높이고, 향후 여러 기관에서 관련 연구 및 사업을 진행할 때 참고할 수 있는 중요한 자료로 이용될 것이다.

3. 향후 연구

1) 개체명 분석의 대상

현재 개체명 분석은 인명, 지명, 기관명 등 고유명 중심으로 이루어지지만, 보통명사 중에도 개체명 분석의 대상이 되는 것이 있다. 특히 7. CIVILIZATION의 상당수가 보통명사에 해당한다. 보통명사는 형태 분석 단계 이후에 일반 사전을 참조하여 그 뜻을 알 수 있는 것이므로 개체명 분석 대상으로 삼을 필요는 없어 보인다. 더욱 간결하고 효율적인 개체명 분석을 위해서는 고유명으로만 그 범위를 한정하는 방안을 검토할 필요가 있다.

2) 개체명 분석 및 개체 연결의 단위

개체명 분석의 단위 결정의 원칙은 최장 단위 태깅이다. 그러나 일부 예외적으로 최소 단위 태깅을 하는 경우가 있다. ‘문재인 대통령’과 같이 인명과 직위명이 함께 나오는 경우 ‘<문재인: PS_NAME> <대통령: CV_POSITION>’과 같이 분리 태깅하는데 개체 연결 시에도 이 개체명 분석 결과를 따라 앞의 <문재인: PS_NAME>만 위키피디아의 ‘문재인’ 페이지에 연결하게 된다. 이 경우는 큰 문제가 되는 것은 아니다. 그러나, ‘문재인 정부’는 개체 연결 시에 문제가 된다. ‘문재인 정부’와 같은 경우 분리 태깅 대상이므로 ‘<문재인: PS_NAME> <정부: OGG_POLITICS>’과 같이 태깅을 하게 된다. 이 구성의 개체 연결 시 <문재인: PS_NAME>은 위키피디아의 ‘문재인’ 페이지에, <정부: OGG_POLITICS>는 위키피디아 보통명사 ‘정부’ 페이지에 연결하게 된다. 위키피디아에 ‘문재인 정부’라는 페이지가 존재하는데도 연결할 수 없는 것이다. 이러한 문제를 해결하기 위해서는 두 가지 방법 중 하나를 택해야 한다. 하나는 개체명 분석에서부터 명사구가 하나의 개체를 가리키는 경우 최장 단위 태깅을 하는 것이다. 그러면 개체명 분석 단위와 개체 연결 단위가 일치하면서도 개체 연결에 문제가 없게 된다. 다른 하나는 개체명 분석은 분리해서 하되 개체 연결 시 분석 단위를 조정하는 것이다. 이 경우 개체 연결은 개체명 분석의 단위를 따른다는 대원칙에 예외에 생기게 되며 이러한 경우에 대한 구체적인 연결 지침을 일일이 제시해야 한다. 실제로 ETRI에서 작성한 ‘Entity Linking 가이드라인 초안’을 보면 문서의 대부분이 이러한 내용으로 채워져 있다. 고유명의 지시대상을 명확히 하기 위해서는 개체명 분석과 개체 연결의 단위에 대한 재검토가 필요하다.

3) 문맥에 따른 개체명 분석 및 개체 연결

개체명 분석과 개체 연결은 문맥에 따라 결정된다. 개체명 분석에서 태깅 범주 결정은 문맥 우선 적용 원칙을 따른다. 개체명의 세분류를 판정할 때 문장이나 문서 전체에서의 의미를 고려하여 태깅하는 것이다. 개체 연결을 할 때도 동일한 개체명 태그로 분석이 되어 있다고 하더라도 지시 대상이 다르면 서로 다른 문서로 태깅하는 것이다. <김정은: PS_NAME>의 경우 개체 연결 시 유명인사만 연결하며, 북한의 최고지도자인 김정은은 ‘김정은’ 페이지로, 배우 김정

은은 ‘김정은(배우)’ 페이지로 연결한다. 이렇게 문맥을 고려하는 태깅은 수작업으로 할 때도 상당한 시간과 노력이 필요하며, 자동 연결을 할 때도 더욱 세심한 주의가 필요하다. 자동 개체명 연결 시 문맥을 고려하여 올바른 페이지로 연결하는 방안에 대한 연구가 필요하다.

4) 보통명사 구 구성의 개체명 분석

보통명사의 개체명 분석에서 구 구성을 최장 단위 태깅할 것인지 최소 단위 태깅할 것인지 결정할 때 <우리말샘> 등재 여부를 따르는 경우가 있다. ‘면바지’는 <우리말샘>에 ‘면-바지’와 같이 한 단어로 등재되어 있어 전체를 <면바지: CV_CLOTHING>과 같이 태깅하고, ‘기모 바지’는 <우리말샘>에 ‘기모^바지’와 같이 구 구성으로 등재되어 있어 전체를 <기모 바지: CV_CLOTHING>와 같이 태깅한다. 그러나 ‘마 바지, 가죽 바지’는 <우리말샘>에 등재되어 있지 않은데 ‘마, 가죽’만 섬유 명칭으로 보아 ‘<마: CV_CLOTHING> 바지, <가죽: CV_CLOTHING> 바지’와 같이 태깅한다. 결과적으로 계열 관계를 이루는 단어들이 서로 다른 방식으로 태깅되는 것이다. 보통명사를 개체명 분석 대상으로 삼지 않으면 문제가 되지 않을 사안이나 개체명 분석 대상에 포함시킨다면 적어도 일관된 처리 방안을 모색해야 할 것이다.

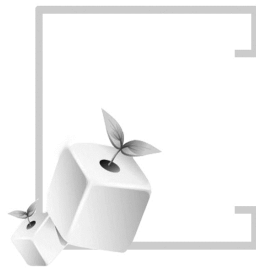
5) 보통명사의 고유명사적 용법

보통명사가 고유명사적 용법을 가지는 경우가 있다. ‘정부는 국민 1인당 25만 원의 재난지원금을 지급하기로 결정했다.’와 같은 신문 기사의 문장에서 ‘정부’는 보통명사이지만, 문맥상 ‘대한민국 정부’ 혹은 ‘문재인 정부’를 의미한다. 이때의 ‘정부’를 <정부: OGG_POLITICS>와 같이 태깅하고 ‘정부’가 특정 국가를 가리킨다고 하더라도 개체명의 언어 표현과 일치하는 위키피디아의 보통명사 ‘정부’ 페이지에 연결한다면 이러한 보통명사의 고유명사적 용법을 표시하지 못하는 한계가 있다. 이러한 용법에 대해 개체명 분석과 개체 연결에서 어떻게 처리할 것인지에 대한 연구가 필요하다.

6) 지식베이스로서 위키피디아의 한계

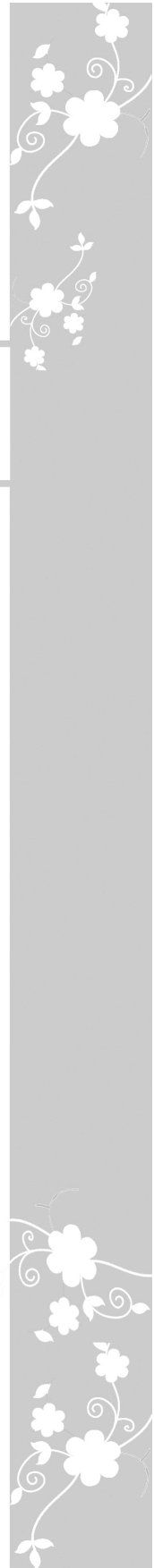
개체 연결은 개체명을 지식베이스인 위키피디아의 해당 페이지에 연결하는

것이다. 그런데 위키피디아는 백과사전, 동적 사전, 오픈 사전이라는 특징을 가지고 있다. 이러한 특성은 개체 연결에 있어서 문제를 야기하기도 한다. 첫째, 위키피디아는 백과사전이기 때문에 여러 개체 유형에 대한 정보가 하나의 문서에 들어 있다. 예를 들어 ‘국립중앙박물관’은 기관으로서의 용법도 가지고, 장소로서의 용법도 가지는데 개체명 분석에서는 이 두 개체명이 서로 다른 태그로 분석되지만, 개체 연결 시에는 하나의 위키피디아 문서에 연결된다. 둘째, 위키피디아는 동적 사전이기 때문에 표제어의 등재 여부와 기술 내용이 실시간으로 달라진다. 개체 연결 결과물을 활용하기 위해서는 어떤 날짜의 위키피디아 페이지에 연결한 것인지에 대한 정보가 부가적으로 필요하다. 셋째, 위키피디아는 오픈 사전이기 때문에 누구나 작성과 편집에 참여할 수 있다. 이로 인해 표제어의 등재 여부와 내용의 기술 방식이 고르지 않다. 때로는 불완전하고 부정확한 정보가 실려 있기도 하다. 위와 같은 문제들 때문에 위키피디아 기반 개체 연결 결과물을 정보 검색 시스템이나 자동 응답 시스템에서 활용할 때 부정확한 정보를 인출할 수 있는 위험이 존재한다. 따라서 위키피디아를 지식베이스로 삼아 개체 연결을 할 때나 개체 연결 결과물을 활용할 때에는 이러한 점에 유의해야 하며 이를 보완하기 위한 연구도 수행해야 할 것이다.



부록

2021 개체명 분석 지침



2021 개체명 분석 지침

ver. 2.1.

2021 개체명 분석 지침

ver. 2.1.

2022.03.04.

2021 개체명 분석 지침

ver. 2.1.

2021 개체명 분석 지침

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. 우선 순위 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
- 2.3. 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

3.6. TR_OTHERS

3.6.1. 정의 및 설명

3.6.2. 분석 대상

3.6.3. 예문

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
- 5.10. 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
- 6.8. 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
 - 7.1.4. 주의 사항
- 7.2. CV_TRIBE
 - 7.2.1. 정의 및 설명
 - 7.2.2. 분석 대상
 - 7.2.3. 예문
- 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
 - 7.6.4. 주의 사항
- 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
- 7.12. 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
 - 8.3.1. 정의 및 설명
 - 8.3.2. 분석 대상
 - 8.3.3. 예문
 - 8.3.4. 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
- 11.5. 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
 - 12.1.4. 주의 사항
- 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

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.
- 15가지의 대분류는 각각 여러 개의 하위범주로 세분화되어 있는데, 목적에 따라 대분류 안의 세부분류를 활용하면 된다. 이 지침은 일종의 가이드라인의 성격을 갖는 것으로, 목적에 따라 필요한 분류를 더 추가하거나 혹은 사용하지 않을 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	

대분류	소분류	세분류
	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
		AFW_OTHER_PRODUCTS
5. ORGANIZATION(OG)	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	
6. LOCATION(LC)	LCP_COUNTRY	
	LCP_PROVINCE	
	LCP_COUNTY	
	LCP_CITY	
	LCP_CAPITALCITY	
	LCG_RIVER	

대분류	소분류	세분류
	LCG_OCEAN	
	LCG_BAY	
	LCG_MOUNTAIN	
	LCG_ISLAND	
	LCG_CONTINENT	
	LC_SPACE	
	LC_OTHERS	
7. CIVILIZATION(CV)	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	
8. DATE(DT)	DT_DURATION	
	DT_DAY	
	DT_WEEK	
	DT_MONTH	

대분류	소분류	세분류
	DT_YEAR	
	DT_SEASON	
	DT_GEOAGE	
	DT_DYNASTY	
	DT_OTHERS	
9. TIME(TI)	TI_DURATION	
	TI_HOUR	
	TI_MINUTE	
	TI_SECOND	
	TI_OTHERS	
10. QUANTITY(QT)	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	
11. EVENT(EV)	EV_ACTIVITY	
	EV_WAR_REVOLUTION	
	EV_SPORTS	
	EV_FESTIVAL	
	EV_OTHERS	

대분류	소분류	세분류
12. ANIMAL(AM)	AM_INSECT	
	AM_BIRD	
	AM_FISH	
	AM_MAMMALIA	
	AM_AMPHIBIA	
	AM_REPTILIA	
	AM_TYPE	
	AM_PART	
	AM_OTHERS	
13. PLANT(PT)	PT_FRUIT	
	PT_FLOWER	
	PT_TREE	
	PT_GRASS	
	PT_TYPE	
	PT_PART	
	PT_OTHERS	
14. MATERIAL(MT)	MT_ELEMENT	
	MT_METAL	
	MT_ROCK	
	MT_CHEMICAL	
15. TERM(TM)	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	

대분류	소분류	세분류
	TMI_PROJECT	
	TMIG_GENRE	
	TM_SPORTS	

분류 명칭	정의 및 예
1. PERSON(PS)	- 인명 및 인물의 별칭
1.1. PS_NAME	- 실존/비실존 인물의 이름, 별칭, 인터넷 사용자명 - 예) 홍선대원군, 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) 뽀뽀, 쿠키, 탄이, 절미, 달리, 김또또, 이몽치
2. STUDY_FIELD(FD)	- 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, 신고전학파
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
3. THEORY(TR)	- 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, 애니미즘, 토테미즘, 샤머니즘
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

분류 명칭	정의 및 예
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) 교통류 이론, 자동초록기법, 텍스트 마이닝 기법
4. ARTIFACTS(AF)	- 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 십층석탑, 에펠탑, 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, 신시사이저
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, KBS2TV
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
5. ORGANIZATION(OG)	- 기관 및 단체
5.1. OGG_ECONOMY	- 경제 관련 기관/단체

분류 명칭	정의 및 예
	- 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, 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
6. LOCATION(LC)	- 지역/장소, 지형/지리

분류 명칭	정의 및 예
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, 랴오둥반도, 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, 금강 휴게소
7. CIVILIZATION(CV)	- 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

분류 명칭	정의 및 예
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/레포트/레저 명칭 - 예) 축구, 야구, 수상스키, 스키, 스케이팅, 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, 맵술령 치킨
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
8. DATE(DT)	- 기간 및 계절, 시기/시대

분류 명칭	정의 및 예
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
9. TIME(TI)	- 시계상으로 나타나는 시/시각, 시간 범위
9.1. TI_DURATION	- 기간 - 예) 6시~9시, 낮, 밤, 점심, 저녁, 오전, 오후, 밤낮
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
10. QUANTITY(QT)	- 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

분류 명칭	정의 및 예
	- 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, 2홉, 20ℓ, 2되, 2만 2000㎤
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
11. EVENT(EV)	- 특정 사건/사고/행사 명칭
11.1. EV_ACTIVITY	- 사회운동 및 선언 명칭 - 예) 공민권운동, 브나로드운동, 노예해방선언, 인권선언, 밀라노칙령
11.2. EV_WAR_REVOLUTION	- 전쟁/혁명, 전투 명칭

분류 명칭	정의 및 예
	- 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
12. ANIMAL(AM)	- 사람을 제외한 짐승
12.1. AM_INSECT	- 곤충 명칭 - 예) 벼메뚜기, 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, 두더지
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, 내항동물, 모악동물, 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)
13. PLANT(PT)	- 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	- 풀이름

분류 명칭	정의 및 예
	- 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
14. MATERIAL(MT)	- 원소 및 금속, 암석/보석, 화학 물질
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
15. TERM(TM)	- 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, 가로무늬, 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

분류 명칭	정의 및 예
15.9. TMI_SW	- IT 소프트웨어 용어 - 예) 윈도우10, 리눅스, 반디집, 알송, 한컴오피스 한글, MS워드, 트로이목마, 자바
15.10. TMI_SITE	- URL 주소 - 예) https://opendict.korean.go.kr/ , https://www.korean.go.kr/
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 사업에서는 위의 대분류 15 항목과 세분류 150 항목의 태그로 개체명을 태깅하고 개체명 분석 말뭉치를 구축한다. 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(보통명사)이 개체명 태깅 대상에 포함된다. 이러한 세부분류명(보통명사)은 <우리말샘> 사전에의 등재 여부에 따라 태깅할 수 있다. 즉, <우리말샘> 사전에 등재되어 있는 보통명사라면 150 항목의 태그로 태깅할 수 있는 개체명 대상 후보가 된다.

두 번째는 '어디부터 어디까지 태깅할 것인가', 즉 개체명 태깅 단위에 관한 문제이다. 단어뿐 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>와 같이 태깅할 수도 있고 <대전: LCP_CITY> <유성구: LCP_COUNTY> <신성동: LCP_COUNTY>와 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 경우에만 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 적정 술어를 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
- 약기
- 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 명사구의 내부 구성요소, 즉 각각의 명사 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'이 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

- <콩순이 말하는 냉장고: AFW_OTHER_PRODUCTS> (O)
- <콩순이: PS_CHARACTER> <말하는 냉장고: AFW_OTHER_PRODUCTS> (X)

- ## 1.2. 최소 단위 태깅(예외)

- 42 -

- <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. 태깅 대상이 여러 개의 단어로 이루어져 있는 경우에는 특정 학문 분야나 이론/법칙과 같은 전문 용어, 예술의 분류나 종류, 여러 장르가 혼합된 새로운 장르, 스포츠나 레저의 명칭 및 종목, 여러 종목이 혼합된 새로운 종목, 친족 관계를 나타내는 명칭, 사람이나 동물이 먹는 음식, 간식의 경우 등과 같이 전체가 하나의 대상을 지시하는 경우가 있다. 이럴 경우에는 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다. 또한 <우리말샘> 사전에 어떻게 등재되어 있는지를 통해 태깅 단위를 결정할 수 있다. <우리말샘> 사전 검색 결과 해당 명사 구성이 '한 단어'로서 등재되어 있는 경우('·', '^'와 같은 기호와 함께 등재되어 있는 경우) '최장 단위 태깅'을 하여 전체를 하나로 태깅한다. <우리말샘>에 '구'로서 등재되어 있거나(띄어쓰기가 되어 있는 경우), 등재되어 있지 않은 경우에도 해당 구가 본래의 의미와 달라지면 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다. 이외에는 '최소 단위 태깅'을 하여 각 단어를 따로 태깅한다. 이러한 사례는 뒤에서 자세히 살필 것이다. 또한 '단어'보다 작은 단위인 복합어(합성어, 파생어) 내부의 구성요소는 개체명에 해당되더라도 따로 태깅하지 않는다. 복합어의 내부 구성요소가 태깅 대상이 되는 경우도 있는데 이러한 사례 또한 뒤에서 자세히 살필 것이다.

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 태깅한다.
 - <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 명사구의 경우, '고모-할머니'와 같이 한 단어로 친족 관계를 나타내는 것과 같이, 구 구성으로 <우리말샘> 사전에 등재되어 있어도 다른 친족 관계 명사에 준해 통합 태깅한다. 구 단위로 사전에 등재되어 있더라도 각각을 따로 태깅하지 않고 '최장 단위 태깅'을 하여 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)
- 직업 명칭을 나타내는 경우, 그 전체가 하나의 대상을 지시하므로 <우리말샘> 사전에 구 구성으로 등재되어 있어도 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다. 또한 직업 명칭은 사회 변화에 따라 새로운 표현이 많이 생성되므로 사전 등재 여부를 떠나 전체가 직업으로 인식되는 표현이라면 '최장 단위 태깅'을 하여 전체를 하나로 태깅한다.
 - 예제
 - 힙합 댄서
 - <힙합 댄서: CV_OCCUPATION> (O)
 - <힙합: CV_ART> <댄서: CV_OCCUPATION> (X)
 - 샹송 가수
 - <샹송 가수: CV_OCCUPATION> (O)
 - <샹송: CV_ART> <가수: CV_OCCUPATION> (X)
 - 파워 블로거
 - <파워 블로거: CV_OCCUPATION> (O)
 - 파워 <블로거: CV_OCCUPATION> (X)
 - '파워 블로거'가 아닌 '블로거'만 나올 경우에는 '블로거'만 CV_OCCUPATION으로 태깅한다. 하지만 '파워 블로거' 전체가 제시될 경우에는 전체를 하나로 태깅한다.
 - 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 드러내지 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)
- 원문의 띄어쓰기 여부와 관계없이 <우리말샘> 사전에 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 본 지침의 세부 내용에 따라 분리 태깅 여부를 결정한다.
 - 예제
 - 소 간, 손가락 뼈: 원문의 띄어쓰기와 관계없이 '소간, 손가락뼈'가 모두 사전에 '-' 기호와 함께 등재되어 있으므로 통합하여 태깅한다.
 - 소 간 → <소 간: TM_CELL_TISSUE_ORGAN>
 - 손가락 뼈 → <손가락 뼈: TM_CELL_TISSUE_ORGAN>
- <우리말샘> 사전에 '^' 기호와 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>
- 구 구성('-', '^'와 같은 기호와 함께 등재되어 있는 경우 포함)의 개체명이 <우리말샘> 사전에 등재되어 있지 않은 경우, 해당 단어가 결합된 구 또는 단어가 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.

2.2. 최소 단위 태깅(예외)

- 실용적인 목적을 위해 명사구의 각 단어의 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, 이 명사구를 단어 단위로 나누어 태깅해도 전체 명사구의 의미가 유지된다면, '최소 단위 태깅'을 한다.

○ 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 경우, 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)
- <우리말샘> 사전에 구 단위로 등재되어 있는 경우, 즉 '-', '^' 기호 없이 띄어쓰기가 되어 있는 경우는 한 단어(복합어)가 아니므로, 명사구 전체는 개체명이 아니어서 태깅하지 않더라도 각 단어가 태깅 대상이라면 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. '부영이카페' 역시 이에 준하여 '부영이'만 AM_BIRD로 태깅한다.
 - 고양이 카페 → <고양이: AM_MAMMALIA> 카페
 - 부영이 카페 → <부영이: AM_BIRD> 카페

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. <우리말샘> 사전에 '^' 기호를 사용하여 등재되어 있는 경우, MT에 한해 '^'를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)한다.

- 마그네슘리본 → <마그네슘: MT_METAL>리본
 - 복합어 내부에서도 금속물적인 의미를 유지하고 있으므로 내부 요소를 태깅한다.
- 산소마스크
 - '산소마스크'는 <우리말샘> 사전에 '산소-마스크'로 등재되어 있다. 따라서 복합어 내부 요소를 태깅하지 않는다.
- 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.

 - 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.

 - 선생님 → <선생님: CV_OCCUPATION>
 - 회장님 → <회장: CV_POSITION>님
- 복합어의 구성요소, 즉 어근이나 접사가 결합된 개체명이 <우리말샘> 사전에 등재되어 있지 않은 경우, 해당 어근 또는 접사가 결합된 단어가 사전에 등재되어 있으면 등재된 것에 준하여 태깅한다.
 - 예제
 - 찌개집: 찌개집의 경우 '꽃집'이 사전에 한 단어로 등재되어 있기 때문에 '찌개'만 별도로 내부 태깅을 하지 않는다.
 - 사이버분쟁 조정사: '사이버분쟁 조정사'는 사전에 등재되어 있지 않으나, '안경사', '상담사' 등은 사전에 등재되어 있다. 따라서 '조정사' 또한 직업의 명칭이 된다. 즉, '사이버분쟁 조정사'는 '안경사', '상담사'와 같이 직업을 나타내는 말이다. 하지만 '사이버분쟁 조정사'를 분리 태깅하면 그 의미가 유지되기 어려우므로, 전체를 'CV_OCCUPATION'으로 태깅한다.

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. 우선 순위 적용(예외)

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

예제	기능 태그 1	기능 태그 2
미니미니, 뽀빠이 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 기능 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 괄호 정보를 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>)
- 개체명과 함께 괄호 정보가 개체명 원어 정보 전체를 나타낼 경우에는 괄호 정보도 따로 태깅한다.
 - 예제
 - 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>
- 주의
 - 단, 원어절 사이에 '음', '이게'와 같이 다른 말이 삽입되어 있는 경우에는 분리하여 태깅한다.
 - 6월부터 이게 8월까지 이어집니다. → <6월부터: DT_OTHERS> 이게 <8월까지: DT_OTHERS> 이어집니다.

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_DURATION
 - 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 삼성: OGG_SPORTS>' (중략) <수원: OGG_SPORTS>이 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 개막. '엑소 수호' 배우 김준면으로 부산 레드카펫 밟는다. → <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 피아노 소나타. 정말 모차르트 너무 어렵다. → 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.

○ 예제

- 보듬 H형 하네스를 구입하기로 했다. H형 하네스를 사면서 리드줄도 함께 구입했다. → <보듬 H형 하네스: AFW_OTHER_PRODUCTS>를 구입하기로 했다. H형 하네스를 사면서 리드줄도 함께 구입했다.
 - 'H형 하네스'는 보통명사로 포괄적으로 지칭하는 것이므로 태깅하지 않는다.
- 서울대학교가 다음 학기 등록금을 동결하기로 했다. 대학교 측은 코로나로 인한 경제 위축과 비대면 수업의 연장 가능성으로 이 같은 결정을 한 것으로 알려졌다. → <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
 - 삼십프로가? → <삼십프로: QT_PERCENTAGE>인가?
 - 러시안데요. → <러시아: LCP_COUNTRY>인데요.
 - 세 겔 → <세 개: QT_COUNT>를

10. 오자나 오타

- 원시 어절에서 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, 캐릭터 명과 반려동물 이름을 포함한 유사 인물 명칭을 대상으로 한다.

분류 명칭	정의 및 설명
1.1. PS_NAME	- 인명 및 인물의 별칭 : 실존 인물의 이름, 비실존 인물(신화 및 종교 인물)의 이름, 별칭
1.2. PS_CHARACTER	- 가상 캐릭터의 이름 - 작품 속 등장인물의 이름
1.3. PS_PET	- 반려동물의 이름

- 주의 사항
 - (태깅 단위 - 접미사 '-이') 받침이 있는 이름에 접미사 '-이'가 붙어 쓰인 경우 이를 포함하여 태깅한다.
 - 영숙이는 집으로 돌아갔다. → <영숙이: PS_NAME>는 집으로 돌아갔다.
 - 우리집 순둥이는 나만 보면 짖는다. → 우리집 <순둥이: PS_PET>는 나만 보면 짖는다.
 - 이때 '-이'는 조사로 실현된 것이 아니기 때문에 포함하여 태깅한다.
 - 김종국이 이번에 대상을 수상했다. → <김종국: PS_NAME>이 이번에 대상을 수상했다.
 - 홍선대원군이 고종을 왕위에 올리기 위해서 → <홍선대원군: PS_NAME>이 고종을 왕위에 올리기 위해서
 - 이때 '이'는 조사이기 때문에 포함하여 태깅하지 않는다.
 - 지영이랑 승원이가 함께 출발했다. → <지영: PS_NAME>이랑 <승원: PS_NAME>가 함께 출발했다.
 - '이랑'은 표준국어대사전에 등재되어 있는 조사이기 때문에 태깅하지 않는다.
 - (태깅 범주 - PS_NAME / PS_CHARACTER / PS_PET) 구문 맥락상 세부 항목 구별이 명확하지 않은 경우에는 PS_NAME을 우선적으로 태깅한다.
 - 내 사랑 미니미니 → 내 사랑 <미니미니: PS_NAME>
 - 우리 뽀뽀이, 엄마가 미안해 → 우리 <뽀뽀이: PS_NAME>, 엄마가 미안해.
 - 나에게 큰 희망과 용기를 준 김지영. 고마워. → 나에게 큰 희망과 용기를 준 <김지영: PS_NAME>. 고마워.

1.1. PS_NAME

1.1.1. 정의 및 설명

- 인명 및 인물의 별칭
 - 실존 인물의 이름
 - 비실존 인물의 이름
 - 별칭

1.1.2. 분석 대상

- 실존 인물의 이름
 - 역사 인물의 이름: 흥선대원군, 해원 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 등장 인물의 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 웨스트매콧
 - 세례명: 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. 윌렌슐레거: PS_NAME>를 선두로 낭만주의 운동이 일어나, <N.그룬트비: PS_NAME> · <안데르센:

PS_NAME> · <키 르케고르: 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 칸영화제 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 <@엇그제: PS_NAME> 이웃님이시네요.

1.1.4. 주의 사항

- (태깅 대상 - 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.
- (태깅 대상 - 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 있기 때문에 PS_NAME으로 태깅하지 않는다.
- (태깅 대상 제외 - 미스, 미즈, 미스터) 영어 이름에 함께 쓰인 미스, 미즈, 미스터는 태깅하지 않는다.
 - 미스 브리트니 → 미스 <브리트니: PS_NAME>
- (태깅 단위 - 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.
 - 이때의 '민'은 특정인이 아닌 가문을 의미하기 때문에 PS_NAME으로 태깅하지 않는다.
- (태깅 단위 - 호+이름) '호+이름' 구성의 경우 결합된 표현이 한 대상을 지칭하므로 전체를 PS_NAME으로 태깅한다. 또한 문서 내에서 호가 단독으로 출현한 경우에도 이를 PS_NAME로 태깅한다. 단, 묘호와 이름 사이에 직위나 직책명이 삽입되어 두 개체명 사이의 명확한 변별이 이루어진 경우에 한정하여 호, 직위명, 본명을 각각 태깅한다.
 - 퇴계 이황은 조선 중기의 학자이자 문신이다. → <퇴계 이황: PS_NAME>은 조선 중기의 학자이자 문신이다.
 - 추사의 글씨를 보고 있노라면 형용할 수 없는 경외감이 싹트곤 한다. → <추사: PS_NAME>의 글씨를 보고 있노라면 형용할 수 없는 경외감이 싹트곤 한다.
 - 정조 대왕 이산은 조선의 제22대 왕이다. → <정조: PS_NAME> <대왕: CV_POSITION> <이산: PS_NAME>은 조선의 제22대 왕이다.
- (태깅 단위 - 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 송송 커플 → 드라마 태양의 후예에서 송중기와 송혜교 커플을 두고 나온 신조어 <송: PS_NAME> <송: PS_NAME> 커플
- (태깅 단위 - 인명+지위, 지위+인명) 다음과 같은 명사구는 지위, 인명, 호칭 따위를 각각 따로 태깅한다.
 - 문재인 대통령 → <문재인: PS_NAME> <대통령: CV_POSITION>
 - 이재용 부회장 → <이재용: PS_NAME> <부회장: CV_POSITION>
 - 국대비 왕씨 → <국대비: CV_POSITION> <왕: PS_NAME> 씨
 - 폐비 윤씨 → <폐비: CV_POSITION> <윤: PS_NAME> 씨
 - 김 박사 → <김: PS_NAME> <박사: CV_POSITION>
- (태깅 단위 - OO황후, OO왕) 묘호나 시호 등이 직위와 함께 쓰여 한 단어로 쓰여 분리가 어려운 경우는 각각 따로 내부 태깅을 하지 않고 한 사람을 지칭하는 말로 볼 수 있으므로 PS_NAME으로 태깅한다.
 - 순정효황후 → <순정효황후: PS_NAME>
 - 명성황후 → <명성황후: PS_NAME>
 - 장수왕 → <장수왕: PS_NAME>
 - 무열왕 → <무열왕: PS_NAME>
- (태깅 단위 - 명사구 인명) 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.
- (태깅 단위 - 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.
- (태깅 단위 - 인명 중첩) 동일 음절 중복으로 두 인명이 이어진 경우, 전체가 특정 인물을 지칭하므로

전체를 하나로 묶어 PS_NAME으로 태깅한다.

- 이명박근혜 → <이명박근혜: PS_NAME>

1.2. PS_CHARACTER

1.2.1. 정의 및 설명

- 가상 캐릭터의 이름
- 작품 속 등장인물의 이름

1.2.2. 분석 대상

- 가상 캐릭터의 이름
 - 만화영화 등장 인물의 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: 펭수, 뿌까, 어피치, 프로도, 무지, 라이언, 브라운, 셸리
 - 공식적인 마스코트의 이름: 블레오, 곰돌이, 마메스케
- 작품 속 등장인물의 이름
 - 소설 등장 인물의 이름: 김지영, 장발장, 해리 포터, 오베, 쇼타
 - 드라마 등장 인물의 이름: 김삼순, 유시진, 박태수, 홍라온, 존 스노우
 - 영화 등장 인물의 이름: 인디애나 존스, 제임스 본드, 피터 파커, 클라크 켄트

1.2.3. 예문

- 가상 캐릭터 이름
 - 이상하게도 요 근래 우리나라는 펭귄 캐릭터가 인기가 많다. <뽀로로: PS_CHARACTER>로 시작해서 <펭수: PS_CHARACTER>까지 이어지는 펭귄 대박 행진을 보라.
 - 라인프렌즈를 대표하는 캐릭터 <브라운: PS_CHARACTER>은 일본에서 특히 인기가 많다.
 - <소닉: PS_CHARACTER>은 고슴도치를 모티브로 만들어졌다.
- 작품 속 등장 인물의 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 스코티 내달 방한한다.
 - 디즈니의 1993년작 알라딘이 실사화를 결정했다. 퍼런 피부와 유쾌한 표정이 트레이드 마크인 <알라

단: PS_CHARACTER>을 과연 누가, 어떻게 소화할 수 있을지 기대된다.

1.2.4. 주의 사항

- (태깅 대상 - 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 역을 맡은 애쉬튼 커처가 팬들을 향해 손을 흔들고 있다. → 영화 '잡스'에서 <스티브 잡스: PS_CHARACTER> 역을 맡은 애쉬튼 커처가 팬들을 향해 손을 흔들고 있다.
 - 영화는 숨을 거두기 직전의 마리 퀴리의 모습을 보이며 시작한다. → 영화는 숨을 거두기 직전의 <마리 퀴리: PS_CHARACTER>의 모습을 보이며 시작한다.
 - 배우 최수종은 "20년 전 출현했던 드라마 한 편에서의 강렬했던 이미지 때문인지 지금까지도 왕건 아저씨라는 호칭을 자주 듣는다"며 웃어 보였다. → 배우 최수종은 "20년 전 출현했던 드라마 한 편에서의 강렬했던 이미지 때문인지 지금까지도 <왕건: PS_CHARACTER> 아저씨라는 호칭을 자주 듣는다"며 웃어 보였다.
- (태깅 대상 - 캐릭터의 별칭) PS_CHARACTER 개체를 캐릭터명이 아닌 또 다른 별칭이나 별명으로 부르는 경우, 이 역시 PS_CHARACTER로 태깅한다.
 - 아니 왜 집어치우시죠? 우리집에도 금빠가 걸어주세요. → 아니 왜 집어치우시죠? 우리집에도 <금빠까: PS_CHARACTER> 걸어 주세요.
 - 가봉이가 가렌 놀리는 말인건 아는데 가렌이랑 무슨 단어가 합쳐진 건가요? → <가봉이: PS_CHARACTER>가 가렌 놀리는 말인건 아는데 가렌이랑 무슨 단어가 합쳐진 건가요?
- (태깅 범주 - PS_CHARACTER / AFA 계열) 구문 맥락상 지칭 대상이 특정한 개인 가상인물이 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.
 - '김지영 보러'의 '김지영'은 영화의 제목을 의미하기 때문에 PS_CHARACTER로 태깅하지 않는다.
 - 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.
 - '해리포터 시리즈'의 '해리포터'는 도서 작품의 이름을 의미하기 때문에 PS_CHARACTER로 태깅하지 않는다.
- (태깅 범주 - PS_CHARACTER / AM 계열 / PT 계열) 우화나 전래동화 속 이야기에서 동물, 식물, 사물 등이 의인화 과정을 통해 사람에 준하는 개체로 판단되는 경우에는 PS_CHARACTER로 태깅한다.
 - 호랑이가 '떡 하나 주면 안 잡아 먹지'라고 말했다. → <호랑이: PS_CHARACTER>가 '떡 하나 주면 안 잡아 먹지'라고 말했다.
 - 어린왕자 속 장미의 명대사는 감명 깊은 대목 중 하나이다. → 어린왕자 속 <장미: PS_CHARACTER>의 명대사는 감명 깊은 대목 중 하나이다.

1.3. PS_PET

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

- (태깅 대상 제외 - 유아어) 주로 어린아이들이 동물을 지칭할 때 사용하는 유아어 '야옹이', '멍멍이', '꿀꿀이' 등은 실제 애완동물의 이름이 아닌 이상 PS_PET이 아닌 AM 계열로 태깅한다.
 - 야옹이는 야옹야옹하고 울어요. → <야옹이: AM_MAMMALIA>는 야옹야옹하고 울어요.
- (태깅 대상 제외 - 동물의 별칭) 온라인상에서 주로 사용되는 개, 고양이 등의 별칭인 '댕댕이', '냥이' 등도 실제 애완동물의 이름이 아닌 이상 PS_PET이 아닌 AM 계열로 태깅한다.
 - 수진이네 댕댕이가 무슨 종이였더라? 리트리버였나? → 수진이네 <댕댕이: AM_MAMMALIA>가 무슨 종이였더라? 리트리버였나?
 - 박신혜, 반려묘와 행복한 일상. 냥이는 집사가 박신혜라 좋겠네. → 박신혜, 반려묘와 행복한 일상. <냥이: AM_MAMMALIA>는 집사가 박신혜라 좋겠네.

2. STUDY_FIELD

- STUDY_FIELD는 학문 분야, 학파 및 유파를 대상으로 한다.

분류 명칭	정의 및 설명
2.1. FD_SCIENCE	- 과학 관련 학문 분야 및 학파 : 순수과학(자연과학), 기술과학(응용과학)
2.2. FD_SOCIAL_SCIENCE	- 사회과학 관련 학문 분야 및 학파 : 인간 사회의 여러 현상을 과학적, 체계적으로 연구하는 모든 경험 학파
2.3.FD_MEDICINE	- 의학 관련 학문 분야 및 학파(분과) : 인간을 질병으로부터 구하고 건강법을 모색하는 학문, 양의학과 한의학을 포함한 각 학문별 세부 학문 명칭
2.4. FD_ART	- 예술 관련 학문 분야 및 학파(유파)
2.5. FD_HUMANITIES	- 인문학 관련 학문 분야 및 학파 : 인간의 삶, 사고 등 인간의 근원 문제에 관해 탐구하는 학문이나 학파
2.6. FD_OTHERS	- FD 계열에 속하지 않는 기타 학문 분야 및 학파

- 주의 사항
 - (태깅 대상 제외 - 학과) 학문 분야로서의 '-학'만 STUDY_FIELD로 태깅한다. 특정 기관의 세부 조직으로서의 '-학과'는 태깅하지 않는다.
 - 물리학을 공부하기 위해서 물리학과에 진학하였다. → <물리학: FD_SCIENCE>을 공부하기 위해서 물리학과에 진학하였다.
 - '물리학과'는 세부 조직에 속하기 때문에 FD_SCIENCE로 태깅하지 않는다.
 - (태깅 대상 제외 - 학파+이론) 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(neoclassical synthesis)의 경우에는 경제적 복지의 달성 수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.
 - '신고전학파 성장론'은 '학파+이론'구성이기 때문에 <신고전학파: FD_SOCIAL_SCIENCE> <성장론: TR_SOCIAL_SCIENCE>으로 태깅하지 않는다.
 - (태깅 대상 제외 - 세분류) 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.

- (태깅 대상 제외 - ○○ 교육) '○○ 교육'의 형태로 쓰인 과목이나 학문명은 태깅하지 않는다. 우리 말샘에 '○○^교육'으로 등록되어 있어 '○○'만 내부 태깅이 불가능하며 '교육'은 태깅 대상이 아니기 때문에 전체 태깅도 하지 않는다.
 - 역사 교육을 소홀히 해서는 안 된다.
 - '역사'만 태깅하지도, '역사 교육'을 태깅하지도 않는다.

2.1. FD_SCIENCE

2.1.1. 정의 및 설명

- 과학 관련 학문 분야 및 과학 학파
 - 순수과학(자연과학)
 - 기술과학(응용과학)

2.1.2. 분석 대상

- 순수과학(자연과학)
 - 물리학, 생명과학, 수학, 지구과학, 화학, 생물물리화학, 소립자물리학, 뉴턴역학, 양자역학, 생물학, 유전학, 응용지질학, 천문학 등
- 기술과학(응용과학)
 - 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

- (태깅 대상 - ○○ 역학) '과학, 물리'와 같은 넓은 범위부터 '뉴턴 역학'과 같은 세부 학문까지 태깅한다. '뉴턴 역학'과 같은 경우는 여러 이론들이 정립된 학문이기 때문에 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 영향을 주었다.
 - '케플러 법칙'과 '뉴턴의 법칙'은 과학 이론이기 때문에 각각 <TR_SCIENCE>로 태깅한다.
- (태깅 대상 - 수의학) '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_SCIENCE로 태깅한다.
 - 인의학과 수의학에는 아직 겹쳐지지 않는 격차가 존재한다. → <인의학: FD_MEDICINE>과 <수의학: FD_SCIENCE>에는 아직 겹쳐지지 않는 격차가 존재한다.
 - '인의학'은 FD_MEDICINE으로 태깅한다.
- (태깅 대상 제외 - 창조 과학, 유사 과학) 학문의 정의에 맞지 않는 대상은 태깅하지 않는다.
 - 창조 과학, 유사 과학, 사이비 과학
- (태깅 단위 - 과목+숫자) '생물1, 생물2'와 같이 번호가 붙은 과목의 경우 '생물'은 FD_SCIENCE로, 번호는 QT_ORDER로 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
 - 인간 사회의 여러 현상을 과학적, 체계적으로 연구하는 모든 경험 과학

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, 뒤르캄학파, 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 <특수교육학: 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 장기동태면에서의 <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 경제현상을 사회학적 방법으로 파악하였고,

2.3. FD_MEDICINE

2.3.1. 정의 및 설명

- 의학 관련 학문 분야 및 학파
 - 인간을 질병으로부터 구하고 건강법을 모색하는 학문
 - 양의학과 한의학을 포함한 각 학문별 세부 학문 명칭

2.3.2. 분석 대상

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

2.3.3. 예문

- 의학 관련 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 연구수단에 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.
- 의학 관련 학파(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

- (태깅 대상 - 의료 분과) 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
- (태깅 대상 제외 - 신체 일부) 신체의 일부에 해당하는 '비뇨기, 피부' 등은 FD_MEDICINE에 해당하지 않는다. 그러나 '비뇨기과, 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.
 - '비뇨 기관, 피부'는 신체의 일부에 해당하기 때문에 FD_MEDICINE으로 태깅하지 않는다.
- (태깅 범주 - FD_MEDICINE / OGG_MEDICINE) 의료 분과의 경우 FD_MEDICINE으로 태깅하고, 구체적인 병원명인 경우 OGG_MEDICINE으로 태깅한다.
 - 목이나 어깨에 불편함이 생기면 정형외과에 방문해 보는 것이 좋다. → 목이나 어깨에 불편함이 생기면 <정형외과: FD_MEDICINE>에 방문해 보는 것이 좋다.
 - 췌장정형외과에서 다리 수술을 했다. → <췌장정형외과: OGG_MEDICINE>에서 다리 수술을 했다.
 - '췌장정형외과'는 특정한 병원명이기 때문에 FD_MEDICINE으로 태깅하지 않는다.

2.4. FD_ART

2.4.1. 정의 및 설명

- 예술 관련 학문 분야 및 학파(유파)
 - 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

- (태깅 대상 제외 - 보통명사) 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.
- (태깅 대상 제외 - 예술 분야) 예술의 한 분야인 경우에는 FD_ART로 태깅하지 않고 CV_ART로 태깅한다.
 - 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.
- (태깅 대상 제외 - 결과물, 작품) 예술의 장르가 아닌 예술 활동의 '결과물(작품)'은 FD_ART로 태깅하지 않는다. 맥락상 '결과물'의 의미로 쓰인 보통명사는 태깅하지 않으며, 결과물의 이름은 AFA 계열로 태깅한다.
 - 사진 700장을 발견했다.
 - 김영하 작가의 신작 소설이 나왔다.

- 맥락상 '사진, 소설'은 예술 활동의 결과물이기 때문에 태깅하지 않는다.
- (태깅 단위 - 학파+예술 분야) '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
 - 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, 스콜라철학, 주기철학, 과학철학, 성리학
 - 스콜라학파, 플라톤학파, 신플로이트파, 유가, 도가, 묵가, 법가, 명가, 음양가, 종횡가, 병가, 기호학파, 영남학파
- 종교학 관련 학문 분야 및 학파
 - 종교학, 기독교학, 가톨릭학, 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)·공 리에서 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

- (태깅 대상 제외 - 종교) 학문의 성격을 가지는 '유교'는 FD_HUMANITIES로 태깅한다. '기독교학, 불교학'과 같이 학문일 경우에는 FD_HUMANITIES로 태깅한다. '기독교, 천주교, 불교' 등 다른 종교는 OGG_RELIGION로 태깅한다.
 - 기독교, 이슬람교, 불교는 세계 3대 종교이다.
 - '기독교, 이슬람교, 불교'는 학문을 뜻하는 것이 아니기 때문에 FD_HUMANITIES로 태깅하지 않는다.
- (태깅 대상 제외 - 조선시대 정치 집단) '기호학파, 영남학파'는 FD_HUMANITIES로 태깅하나 '노론, 소론'과 같은 조선시대 정치 집단은 OGG_POLITICS로 태깅한다.
 - 서인에서 갈라져 나온 노론은 성리학을 신봉하였다. → 서인에서 갈라져 나온 <노론: OGG_POLITICS>은 <성리학: FD_HUMANITIES>을 신봉하였다.
 - '노론'은 정치 집단이기 때문에 FD_HUMANITIES로 태깅하지 않는다.
- (태깅 대상 제외 - 철학 단체) '빈 학단'과 같은 철학 단체는 FD_HUMANITIES로 태깅하지 않는다.
 - 빈 학단의 목표는 철학을 과학적으로 만드는 것이었다.
 - '빈 학단'은 철학 단체에 속하기 때문에 FD_HUMANITIES로 태깅하지 않는다.
- (태깅 단위 - 주의+학파) '○○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>
- (태깅 범주 - FD_HUMANITIES / CV_LANGUAGE) '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.

- 위 문장에서 '영어'는 언어를 뜻하는 것이기 때문에 CV_LANGUAGE로 태깅한다.

2.6. FD_OTHERS

2.6.1. 정의 및 설명

- FD 계열에 속하지 않는 기타 학문 분야 및 학파

2.6.2. 분석 대상

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

2.6.3. 예문

- FD 계열에 속하지 않는 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 대상으로 한다. 대학 전공 서적과 같은 전문적인 이론, 법칙 관련 용어들을 정의 또는 설명하고 있는 경우 TR로 태깅한다.

분류 명칭	정의 및 설명
3.1. TR_SCIENCE	- 과학 관련 이론/법칙/방식/양식/체계/학설 : 순수과학, 기술과학
3.2. TR_SOCIAL_SCIENCE	- 사회과학 이론/법칙/방법/원리/사상 - 인류학 및 문화인류학 사상 - 정치사상
3.3. TR_MEDICINE	- 의학 요법/처방, 의학 진단법
3.4. TR_ART	- 예술 관련 이론/법칙/방식/양식 - 예술 사조 - 촬영/연출 기법
3.5. TR_HUMANITIES	- 철학, 언어, 역사, 종교적 이론/사상
3.6. TR_OTHERS	- TR 계열에 속하지 않는 기타 이론/법칙/원리

- 주의 사항
 - (태깅 대상 제외 - 일반 용어) 일반적인 용어들은 태깅하지 않는다.
 - 각 나라의 사회법칙에 따라 용인될 수 있는 행동이 다르다.
 - '사회법칙'은 일반적인 용어이기 때문에 THEORY로 태깅하지 않는다.
 - (태깅 단위 - 사상/원리/이론/주의) '사상/원리/이론/주의/효과'와 함께 쓰일 경우, '- 사상/원리/이론/주의'를 포함하여 태깅한다.
 - 특수상대성이론 → <특수상대성이론: TR_SCIENCE>
 - (특수 항목 태깅 - 기호 삽입) THEORY가 '사상/원리/이론/주의'와 함께 쓰일 때 기호가 함께 쓰일 수 있다. 이 경우 기호를 포함해서 THEORY로 태깅한다.
 - 멘델에 의해 정리되어 명명된 '멘델'의 법칙 → 멘델에 의해 정리되어 명명된 <'멘델'의 법칙: TR_SCIENCE>

3.1. TR_SCIENCE

3.1.1. 정의 및 설명

- 과학 관련 이론/법칙/방식/양식/체계/학설
 - 순수과학
 - 기술과학

3.1.2. 분석 대상

- 순수과학
 - 특수상대성이론, 우열의 법칙, 양자론, 멘델법칙, 드무아브르의 정리, 상대성원리
 - 21세기형수학공식, 이진법, 3원질설
- 기술과학
 - 원심분리법, 기체확산법, 열확산법, 컨베이어 라인(conveyer line) 시스템
 - 이온주입법, 공랭식, 수랭식
 - 인공지능, LTE-A, 아이맥스

3.1.3. 예문

- 순수과학
 - 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 물리 법칙이 같은 형식으로 기술되도록 정식화(定式化)되어 있다.
- 기술과학
 - 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.
 - <인공지능: TR_SCIENCE>을 이용한 스피커가 인기를 얻고 있다.

3.1.4. 주의 사항

- (태깅 대상 - 특정 인물+이론/효과/현상) 어떤 이론이나 현상을 특정 인물이 발견하거나 정리하여 그 인물의 이름이 붙은 경우는 TR_SCIENCE로 태깅한다.
- 도플러가 처음 제안한 도플러 효과는 전자기파 파장과 관련된 현상이다. → <도플러: PS_NAME>가 처음 제안한 <도플러 효과: TR_SCIENCE>는 전자기파 파장과 관련된 현상이다.
 - 사람 이름으로 쓰인 '도플러'가 아닌 과학 이론의 '도플러 효과'는 TR_SCIENCE로 태깅한다.
- (태깅 대상 - 가설) 입증된 주장이나 이론이 아닌 가설일 경우, 해당 가설을 주장한 사람과 내용이 명

확하다면 TR_SCIENCE로 태깅한다.

- 프랙탈 우주론은 아직 많은 사람들에게 받아들여지지 않았다. → <프랙탈 우주론: TR_SCIENCE>은 아직 많은 사람들에게 받아들여지지 않았다.
- (태깅 대상 제외 - 수학 용어) '2진수', '자연수', '약수' 등은 과학 이론이 아니므로 TR_SCIENCE로 태깅하지 않는다.
 - 컴퓨터는 2진수로 정보를 변환한다.
 - 2진수는 특정한 과학 이론이 아니기 때문에 TR_SCIENCE로 태깅하지 않는다.
- (태깅 대상 제외 - 포괄적인 개념, 현상) 고유의 명칭을 갖고 있는 특정 과학 이론만이 태깅 대상이며, 여러 구체적인 이론을 포함할 수 있는 포괄적인 개념의 이론/원리, 단순한 작용, 현상은 태깅하지 않는다.
 - 형광작용, 중력, 빙식작용, 관성력, 사진작용, 이온화작용, 파쇄작용, 연마작용, 자기장, 적외선

3.2. TR_SOCIAL_SCIENCE

3.2.1. 정의 및 설명

- 사회과학 이론/법칙/방법/원리/사상
- 인류학 및 문화인류학 사상
- 정치사상

3.2.2. 분석 대상

- 사회과학 이론/법칙/방법/원리/사상
 - 무작위추출법, 노동가치설, 대화법
 - 하인리히의 법칙, 수요와 공급의 법칙
 - 삼포식 농업
 - 법치주의, 평등주의, 관료주의, 민족주의, 자본주의, 개인주의
- 인류학 및 문화인류학 사상
 - 애니미즘, 토테미즘, 샤머니즘
- 정치 사상
 - 민주주의, 공산주의, 다원적 민주주의, 사회주의, 자유주의

3.2.3. 예문

- 사회과학 이론/법칙/방법/원리/사상
 - <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>
 - 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,
- 인류학 및 문화인류학 용어
 - 토착종교인 <샤머니즘: TR_SOCIAL_SCIENCE>의 고원교를 몰아내는 투쟁과 그 승리를 극명하게 표현한다.
 - <애니미즘: TR_SOCIAL_SCIENCE>은 영국의 타일러 인류학자가 처음 사용한 단어이다.
 - <토테미즘: TR_SOCIAL_SCIENCE>에서는 자신이 속한 집단과 특수한 관계가 있는 동물이나 식물을 신성시 여긴다.
- 정치사상
 - <사회주의: TR_SOCIAL_SCIENCE>란 말이 <개인주의: TR_SOCIAL_SCIENCE>의 반대말로서 르루에 의하여 만들어졌음을 알 수 있다.
 - 개방된 시장경제와 <다원적 민주주의: TR_SOCIAL_SCIENCE>라는 가치관을 공유하는 국가간 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,

3.2.4. 주의 사항

- (태깅 대상 - ○○ 콤플렉스) '로리타 콤플렉스', '오이디푸스 콤플렉스' 등의 콤플렉스 이름은 심리학의 이론이므로 TR_SOCIAL_SCIENCE로 태깅한다.
 - 그는 자신이 왜 오이디푸스 콤플렉스에 시달리는지 알지 못했다. → 그는 자신이 왜 <오이디푸스 콤플렉스: TR_SOCIAL_SCIENCE>에 시달리는지 알지 못했다.
- (태깅 단위 - 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.

3.3. TR_MEDICINE

3.3.1. 정의 및 설명

- 의학 요법/처방, 의학 진단법

3.3.2. 분석 대상

- 의학 요법/처방, 의학 진단법
 - 팔성혈 자세교정법, 인공호흡, 요혈자극법, 뜸, 바이오피드백
 - 발열요법, 견인요법, 보법, 대기요법, 공기요법
 - 수지력 테스트, X 선검사
 - 맥진, 내시경검사, 혈액검사, 검안, 문진, 예방접종

3.3.3 예문

- 의학 요법/처방, 의학 진단법
 - <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.
 - 촉각으로 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

- 예술 관련 이론/법칙/방식/양식
 - 에로티시즘, 원근법, 대기원근법, 선원근법, 구름법, 물골법
 - 대위법, 데카당스, 문자기보법, 환유법, 제유법
 - 몽타주, 오마주, 추사체, 데칼코마니, 액션페인팅
- 예술사조
 - 고전주의, 낭만주의, 초현실주의, 표현주의
- 촬영/연출 기법
 - 익스트림 롱 샷, 풀 피겨 샷, 클로즈업, 루즈샷, 패닝, 틸트

3.4.3 예문

- 예술 관련 이론/법칙/방식/양식
 - <고딕양식: TR_ART>의 발달로 교회의 외관은 특색이 강해졌고
 - <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 고전양식을 성립하게 한다.
 - 영화에서 <에로티시즘: TR_ART>은 존재의 문제에 접근하려는 시도를 보였으나 때로는 성행위를 노골적으로 묘사한 포르노그래피에 불과한 경우도 많았다.
 - <대위법: TR_ART>에서는 각 성부가 명료하게 식별할 수 있는 선율적 독립성을 지니며,
 - 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.
- 예술사조
 - <고전주의: TR_ART>의 부활이야말로
 - 이후 그들의 예술적 경향을 <데카당스: TR_ART>라고 평하였다.
- 촬영/연출 기법
 - 움직임이 많은 대상을 촬영할 때는 <패닝: TR_ART>을 자주 사용한다.
 - <클로즈업: TR_ART>을 이용한 이 장면으로 미세한 표정 변화를 확인할 수 있다.

3.4.4. 주의 사항

- (태깅 대상 - 예술운동) '다다이즘' 등의 예술운동은 일종의 예술사조로 고려해서 TR_ART로 태깅한다.
 - 다다이즘의 예술운동은 잘 알려지지 않았지만 → <다다이즘: TR_ART>의 예술운동은 잘 알려지지 않았지만
- (태깅 대상 제외 - 르네상스) '르네상스'는 예술운동 외의 학문, 사상 등까지 다루고 있으므로 예술사조로 보기 힘들다. '르네상스'는 일종의 문화혁신운동이므로 'CV_CULTURE'로 태깅한다

- 르네상스는 14세기 중반에 인문주의를 바탕으로 → <르네상스: CV_CULTURE>는 14세기 중반에 인문주의를 바탕으로
 - '르네상스'는 예술운동에 국한되는 것이 아니기 때문에 TR_ART로 태깅하지 않고 CV_CULTURE로 태깅한다.

3.5. TR_HUMANITIES

3.5.1. 정의 및 설명

- 철학, 언어, 역사, 종교적 이론/사상

3.5.2. 분석 대상

- 철학, 언어, 역사, 종교적 이론/사상
 - 조작주의, 주지주의, 구조주의, 토마스주의, 경험주의, 합리주의, 공리주의, 현실주의
 - 니힐리즘, 이데아설, 인식론, 노장사상, 이기론, 존재론, 관념론

3.5.3 예문

- 철학, 언어, 역사, 종교적 이론/사상
 - 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 감각에 의한 것이 아니라 지성에 의해서 생긴다고 보는 합리론도 넓은 뜻의 <주지주의: TR_HUMANITIES>이다.

3.6. TR_OTHERS

3.6.1. 정의 및 설명

- TR 계열에 속하지 않는 기타 이론/법칙/원리

3.6.2. 분석 대상

- 교통류 이론, 자동초록기법, 텍스트 마이닝 기법

3.6.3. 예문

- 논문 작성자들은 <자동초록기법: TR_OTHERS>으로 필수 정보를 선택적으로 제시할 수 있게 되었다.
- 도시 개발을 위해서는 <교통류 이론: TR_OTHERS>에 대한 이해가 선행되어야 한다.
- <텍스트 마이닝 기법: TR_OTHERS>은 어휘나 정보 추출뿐만 아니라 문장 패턴까지 분석 대상으로 삼는다.

4. ARTIFACTS

- ARTIFACTS는 사람에 의해 창조된 인공물을 대상으로 한다. ARTIFACTS 세부 범주에 해당되지 않는 인공물은 태깅하지 않는다.

분류 명칭	정의 및 설명
4.1. AF_BUILDING	- 건축물/토목건설물 : 건축, 다리, 댐, 보, 기타 건설물의 이름
4.2. AF_CULTURAL_ASSET	- 건축물, 고분, 사찰 등의 문화재 - 문화재로 지정되어 있지 않은 조선시대까지 만들어진 절과 무덤 - 기타 해외 문화재
4.3. AF_ROAD	- 도로, 철로, 수로 등 길의 명칭 - 관광지 거리/골목명
4.4. AF_TRANSPORT	- 교통수단/자동차/선박 모델 및 유형 - 운송수단, 놀이기구
4.5. AF_MUSICAL_INSTRUMENT	- 악기 명칭 : 현악기, 관악기, 타악기, 전자악기 - 적정 술어 : '~을 연주하다/켜다/타다/치다/불다' 등
4.6. AF_WEAPON	- 무기 명칭 : 도검류, 총포류, 폭탄류, 기타 무기류
4.7. AF_ART_WORKS	- 작품명
4.7.1. AFA_DOCUMENT	- 도서, 문학작품, 기타 문서의 명칭
4.7.2. AFA_PERFORMANCE	- 춤/무용 작품명, 연극/공연/가극/뮤지컬 제목
4.7.3. AFA_VIDEO	- 영화 작품명, TV 프로그램 이름, TV/라디오 채널명
4.7.4. AFA_ART_CRAFT	- 미술 작품명 : 그림, 조각 등의 제목, 기타 미술품의 제목
4.7.5. AFA_MUSIC	- 음악 작품명 : 노래의 제목, 오페라/뮤직비디오 작품명
4.8. AF_WARES	- 판매되는 상품 및 제품
4.8.1. AFW_SERVICE_PRODUCTS	- 서비스업의 상품 명칭 : 금융/여행/교육 상품명
4.8.2. AFW_OTHER_PRODUCTS	- 식품과 음료를 제외한 기타 공산품의 상품명

4.1. AF_BUILDING

4.1.1. 정의 및 설명

- 건축물/토목건설물
 - 건물
 - 다리

- 댐, 보
- 기타 건설물

4.1.2. 분석 대상

- 건물 이름
 - 한화생명 63빌딩, 롯데월드타워, Two IFC, 서울종합운동장 야구장, 잠실실내체육관, 대우트럼프월드 I, 은마아파트, 백악관, 파리 노트르담 대성당 등
- 다리 이름
 - 성수대교, 한강철교, 풍네프, 골든 게이트 브리지 등
- 댐, 보 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
 - 헐리우드 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

- (태깅 대상 - 아파트 명, 아파트 동) '아파트 명', '아파트 명+OO동', 아파트 명 없이 'OO동'만 쓰인 경우, AF_BUILDING으로 태깅한다.
 - 대치동 은마아파트 → 대치동 <은마아파트: AF_BUILDING>
 - 대치동 은마아파트 101동 → 대치동 <은마아파트 101동: AF_BUILDING>
 - 301동에 사는 사람들 → <301동: AF_BUILDING>에 사는 사람들
- (태깅 대상 - 정식 명칭) 건축물이 지명이나 상위 조직과 함께 쓰인 경우, 전체가 정식 명칭이면 분리

하지 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.
- 서울 올림픽공원 체조경기장 → <서울: LCP_CAPITALCITY> <올림픽공원: LC_OTHERS> 체조경기장
 - 일상적으로 많이 쓰이더라도 정식 명칭이 아니면 전체를 하나의 명칭으로 태깅하지 않는다.
- (태깅 범주 - AF_BUILDING / OGG 계열) 어떠한 단어가 건축물(AF)이면서 기관(OG)의 의미를 지닐 경우, '~이 보인다/붕괴하다/흔들리다, ~을 짓다/만들다'와 함께 쓰이는 등 문맥에 의해 건축물의 의미로 해석되는 경우만 AF_BUILDING으로 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>
- (태깅 범주 - AF_BUILDING / OGG_SPORTS) 골프장, 수영장 등과 같은 운동 시설이 건물의 의미일 때는 AF_BUILDING으로 태깅하지만 기관의 의미로 쓰일 때에는 OGG_SPORTS로 태깅한다.
- 서강대교를 지나면 여의도 한강 공원 수영장이 보인다. → 서강대교를 지나면 <여의도 한강 공원 수영장: AF_BUILDING>이 보인다.
- 여의도 한강 공원 수영장은 많은 피서객들을 유치하기 위해 시설 정비에 들어갔다. → <여의도 한강 공원 수영장: OGG_SPORTS>은 많은 피서객들을 유치하기 위해 시설 정비에 들어갔다.
 - 위 문장에서 '여의도 한강 공원 수영장'은 건물이 아닌 기관의 의미이기 때문에 OGG_SPORTS로 태깅한다.
- (태깅 범주 - 경포대) '경포대'는 문맥에 따라 해변, 건물 등으로 모두 태깅 가능하다. '세우다, 만들다' 등의 술어와 사용되면 AF_BUILDING으로 태깅하고 '붐비다, 놀러 가다' 등의 술어와 함께 사용되면 LC_OTHERS로 태깅한다. 술어를 참고하되 술어의 다의성에 주의하여 태깅하며, 문맥으로 판단이 어려울 경우 LC_OTHERS로 우선 태깅한다.
- 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?
 - 문맥상 '누각'을 지칭하므로 AF_BUILDING으로 태깅한다.
- 동해안의 대표적인 해변인 경포대 일출은 사실 어느 지점에서 봐도 동일해요. → 동해안의 대표적인 해변인 <경포대: LC_OTHERS> 일출은 사실 어느 지점에서 봐도 동일해요.
 - 문맥상 '해변'을 지칭하므로 LC_OTHERS로 태깅한다.
- (태깅 범주 - AF_BUILDING / AF_CULTURAL_ASSET) ARTIFACTS에 해당하는 대상이 AF_BUILDING인지 AF_CULTURAL_ASSET인지 문맥으로 판단할 수 없는 경우에는 본유적 의미인 AF_BUILDING을 우선하여 태깅한다.
- 관덕정 앞에서 찍은 가족사진 → <관덕정: AF_BUILDING> 앞에서 찍은 가족사진
 - 위 '관덕정'은 다른 설명이 없어 문화재인지 판단이 불가하므로 AF_BUILDING으로 태깅한다.
- 제주시에 있는 관덕정 → 제주시에 있는 <관덕정: AF_CULTURAL_ASSET>
 - 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.

- (태깅 범주 - AF_BUILDING / OGG_ART) 예술 기관은 문맥에 따라 AF_BUILDING과 OGG_ART로 태깅이 가능한데 건축물이 아닌 기관의 의미로 사용되면 OGG_ART로 태깅한다. '~이 열린다'와 같은 서술어와 어울리면 AF_BUILDING로 태깅하고 '~를 운영하다, ~를 경영하다, ~이 망하다'와 같은 서술어와 어울리면 OGG_ART로 태깅한다.
 - 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.
 - 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.
- (태깅 범주 - AF_BUILDING / OGG 계열 / LC_OTHERS) 건물이 AF_BUILDING, OGG 계열, LC_OTHERS로 중복 태깅이 가능할 경우 문맥과 술어를 고려하여 태깅한다. 단, OGG 계열이 나왔을 때 개별 상점의 경우 '~에 가다, ~에서 만나다' 등의 술어가 나왔다고 하더라도 OGG 계열로 태깅한다. 일정한 구역에 여러 상점이 한곳에 모여 있는 복합문화공간, 쇼핑몰의 경우에만 한정하여 LC_OTHERS로 태깅한다. '서면롯데백화점, 현대백화점 판교'와 같이 특정한 'OO백화점'이나 '홍대/이대' 등 특정한 'OO대(학교)'와 같은 대상은 '~이 보이다/흔들리다, ~을 짓다/만들다'와 함께 쓰이며 건축물의 의미로 해석되는 경우 AF_BUILDING으로 태깅한다. '~에 다니다, ~을 경영하다, ~이 흥하다, ~이 잘 되다, ~을 차리다'와 같은 술어와 함께 쓰이면 OGG로 태깅하고 '~에서 만나다/구입하다/팔다/사다' 등과 함께 사용되어 '장소/위치'를 의미하는 경우 LC_OTHERS로 태깅한다.
 - 명동에 있는 <롯데백화점: AF_BUILDING>은 1979년에 지어졌다.
 - '기관'이 아닌 '건축물'의 개념으로 사용되었으므로 AF_BUILDING으로 태깅한다
 - 그는 광양제철고-중앙대를 거쳐 안산 그리너스에서 프로에 데뷔했다. → 그는 <광양제철고: OGG_EDUCATION>-<중앙대: OGG_EDUCATION>를 거쳐 안산 그리너스에서 프로에 데뷔했다.
 - 장소/위치가 아닌 졸업한 기관을 뜻하므로 OGG_EDUCATION으로 태깅한다.
 - 어제 홍대에서 산 핑크색 거울 니트 → 어제 <홍대: LC_OTHERS>에서 산 핑크색 거울 니트
 - 서강대 근처에 있는 카페에서 → <서강대: LC_OTHERS> 근처에 있는 카페에서
 - 위 예문들의 '홍대, 서강대'는 학교 일대 지역을 뜻하므로 LC_OTHERS로 태깅한다.
 - 퇴근 후에 다이소에 갔다. → 퇴근 후에 <다이소: OGG_ECONOMY>에 갔다.
 - 스타벅스에 가서 산 텀블러. → <스타벅스 : OGG_FOOD>에 가서 산 텀블러.
 - CGV에 가서 봤던 영화가 → <CGV : OGG_ART>에 가서 봤던 영화가
 - '다이소, 스타벅스, CGV'가 '-에 가다'의 서술어와 쓰여 LC로 볼 수 있으나 OGG 계열의 개별 상점이므로 OGG로 태깅한다.

4.2. AF_CULTURAL_ASSET

4.2.1. 정의 및 설명

- 건축물, 고분, 사찰 등의 문화재

- 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 석등 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

- (태깅 대상 - 건축 문화재) 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.

- (태깅 대상 - 자연유산) 동식물 등의 자연유산은 AM 계열, PT 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>
- (태깅 대상 제외 - 인간문화재) 인간문화재는 '인간문화재 000'과 같이 표현되므로 무형문화재를 보유한 사람 이름을 PS_NAME으로 태깅한다.
 - 인간문화재 공옥진 → 인간문화재 <공옥진: PS_NAME>
- (태깅 대상 제외 - 보물, 국보, 유산) '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>
- (태깅 단위 - 서책, 함) 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>
- (태깅 범주 - AF_CULTURAL_ASSET / AF_BUILDING) ARTIFACTS에 해당하는 대상이 AF_CULTURAL_ASSET인지 AF_BUILDING인지 문맥으로 판단할 수 없는 경우에는 본유적 의미인 AF_BUILDING을 우선하여 태깅한다.
 - 제주시에 있는 관덕정에 방문하였다. → 제주시에 있는 <관덕정: AF_CULTURAL_ASSET>에 방문하였다.
 - 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. 위 문장의 '관덕정'은 다른 설명이 없어 문화재인지 판단이 불가하므로 AF_BUILDING으로 태깅한다.

4.3. AF_ROAD

4.3.1. 정의 및 설명

- 도로, 철로, 수로 등 길의 명칭
 - 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

- (태깅 대상 - OO거리, OO길) '명동거리', '덕수궁돌담길', '경리단길' 등 길에 위치한 다른 즐길 거리/길을 포함하는 명칭도 AF_ROAD로 태깅한다.
 - 요즘 젊은이들 사이에서 <경리단길: AF_ROAD> 투어가 핫하다.
- (태깅 대상 제외 - 불분명한 도로) 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 않기 때문에 '북한산 등산로'는 AF_ROAD로 태깅하지 않는다.

- (태깅 단위 - 국도) '국도1호선, 44번 국도' 등 고유의 번호를 가진 국도는 통합하여 태깅한다.
 - 11번 국도 → <11번 국도: AF_ROAD>
- (태깅 단위 - 도로명 + n길, n로) 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>
- (태깅 범주 - AF_ROAD / LC_OTHERS) ○○사거리, ○○로터리 등 길의 명칭이 맥락에 따라 '~에 있다/위치하다' 등과 함께 쓰여 특정 위치를 의미하는 경우는 LC_OTHERS로 태깅한다.
 - 미아사거리가 가장 막히는 구간이다. → <미아사거리: LC_OTHERS>가 가장 막히는 구간이다.
 - 미아사거리에 있는 현대백화점에서 만나자. → <미아사거리: LC_OTHERS>에 있는 현대백화점에서 만나자.

4.4. AF_TRANSPORT

4.4.1. 정의 및 설명

- 운송 수단
- 놀이기구

4.4.2. 분석 대상

- 운송 수단
 - 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
- 놀이기구
 - 롤러코스터, 바이킹, 범퍼카 등

4.4.3. 예문

- 운송 수단의 하위 유형
 - 휴갯길 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 해외선 <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

- (태깅 대상 - 고유명사) '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.
- (태깅 대상 - 함선) 무기를 갖춘 함선은 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.
- (태깅 대상 제외 - 인공위성) 운송 수단이 아닌 인공위성은 TMI_HW로 태깅한다.
 - 과학기술 3호 → <과학기술 3호: TMI_HW>
 - '과학기술 3호'는 인공위성의 고유명이기 때문에 AF_TRANSPORT로 태깅하지 않는다.

4.5. AF_MUSICAL_INSTRUMENT

4.5.1. 정의 및 설명

- 악기 명칭
 - 현악기
 - 관악기
 - 타악기
 - 전자악기

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
- 전자악기
 - 신시사이저, 키보드, 전자드럼, 런치패드 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.
 - <꽁과리: AF_MUSICAL_INSTRUMENT>가 종묘제례악(宗廟祭禮樂)에 쓰일 때는 다른 이름으로 불린다.
- 전자악기
 - 다양한 악기의 소리를 표현할 수 있는 <신시사이저: AF_MUSICAL_INSTRUMENT>
 - 자신의 비트를 <런치패드: AF_MUSICAL_INSTRUMENT>로 찍어서 영상을 올리곤 한다.

4.5.4. 주의 사항

- (태깅 대상 - 전자악기의 종류) 악기의 경우 '~을 연주하다/켜다/타다/치다/불다' 등의 적정 술어와 함께 쓰이는 경우가 많다. 그러나 전자악기의 경우, '~ 소리를 만들다/찍다/표현하다' 등의 적정 술어가 쓰일 수 있다.
 - 그 노래는 배우가 드라마에서 직접 기타를 치며 노래를 불러 유명해졌다. → 그 노래는 배우가 드라마에서 직접 <기타: AF_MUSICAL_INSTRUMENT>를 치며 노래를 불러 유명해졌다.
 - 신시사이저로 만들어 낸 소리가 제법 진짜 악기 같다. → <신시사이저: AF_MUSICAL_INSTRUMENT>로 만들어 낸 소리가 제법 진짜 악기 같다.
- (태깅 대상 - 고유명사) 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
 - 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

- (태깅 대상 - 고유명) 'k-2', 'M16A1'처럼 무기의 고유 명칭(상품명)이 쓰인 경우에도 AF_WEAPON으로 태깅한다.
 - 한국형 소총 K-2는 대한민국 육군의 기본 화력을 구성한다. → 한국형 <소총: AF_WEAPON> <K-2: AF_WEAPON>는 대한민국 육군의 기본 화력을 구성한다.
- (태깅 대상 - 문화재) '승자총통', '비격진천뢰'와 같이 조선시대 이전에 제작되어 문화재로 등록된 무기

의 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.
- (태깅 대상 - 특정 문맥에서의 무기) 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

- 작품명
- 주의 사항
 - (태깅 대상 제외 - 코너/장/막 등의 명칭) 프로그램 속 코너나 작품 속에 포함되어 있는 장이나 막 등의 명칭은 그 프로그램이나 작품의 일부로 보아 태깅하지 않는다.
 - 코미디빅리그 사이코러스가 큰 인기를 얻고 있다. → <코미디빅리그 : AFA_VIDEO>의 사이코러스가 큰 인기를 얻고 있다.
 - 데미안에서 가장 기억에 남는 장면은 '새는 알에서 나오려고 투쟁한다'이다. → <데미안: AFA_DOCUMENT>에서 가장 기억에 남는 '새는 알에서 나오려고 투쟁한다'이다.

4.7.1. AFA_DOCUMENT

4.7.1.1. 정의 및 설명

- 도서, 문학 작품, 기타 문서의 명칭

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 만들기가 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

- (태깅 대상 - 훈민정음) 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 AFA_DOCUMENT로 태깅하지 않는다.
- (태깅 대상 - 웹툰명, 만화책명) 웹툰이나 만화책의 작품명은 AFA_DOCUMENT로 태깅한다.
 - 90년대에 태어난 아이들은 드래곤볼, 슬램덩크 등 일본 만화책에 익숙하다. → 90년대에 태어난 아이들은 <드래곤볼: AFA_DOCUMENT>, <슬램덩크: AFA_DOCUMENT> 등 일본 만화책에 익숙하다.
- (태깅 대상 제외 - 애니메이션 작품명) 애니메이션은 AFA_VIDEO로 태깅한다.
 - TV 앞에는 겨울왕국을 보려는 아이들이 모여 들었다. → TV 앞에는 <겨울왕국: AFA_VIDEO>을 보려는 아이들이 모여 들었다.
- (태깅 범주 - AFA_DOCUMENT / AFA_PERFORMANCE) 대상이 시나리오를 뜻하는 경우에는 AFA_DOCUMENT로 태깅하고 연극 작품명이 명확한 경우에만 AFA_PERFORMANCE로 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.
 - 위 문장에서 '로미오와 줄리엣'은 연극 작품명의 의미이기 때문에 AFA_DOCUMENT로 태깅하지 않는다.

4.7.2. AFA_PERFORMANCE

4.7.2.1. 정의 및 설명

- 춤, 무용, 연극 등 공연의 작품명

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

- (태깅 대상 제외 - 춤의 종류) 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.
- (태깅 대상 제외 - 오페라) 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

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, SBS러브FM, 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
 - 한국관광공사에서 한국을 홍보하는 영상 <Feel the Rhythm of Korea: SEOUL: AFA_VIDEO>를 공개했다.
- 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

- (태깅 대상 - 줄임말) 작품명의 일부가 절단된 줄임말의 경우, 맥락을 통해 확인이 가능하면 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?
- (태깅 대상 - 뮤직비디오 제목) 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.
- (태깅 단위 - 시즌, 시리즈) 프로그램명과 시리즈나 시즌이 함께 있는 경우는 포함하여 태깅한다.
 - 슬기로운 의사생활 시즌 1 → <슬기로운 의사생활 시즌 1: AFA_VIDEO>
 - 슬기로운 의사생활 시즌 1에 이어 시즌 2에서는 → <슬기로운 의사생활 시즌 1: AFA_VIDEO>에 이어 <시즌 2: AFA_VIDEO>
 - 문맥으로 어떤 프로그램명인지 알 수 있는 경우에는 AFA_VIDEO로 태깅한다.
 - 시즌 1에 나온 멤버들과 다르게 → <시즌 1: QT_ORDER>에 나온 멤버들과 다르게
 - 단, 프로그램명이 쓰이지 않고 단독으로 쓰인 '시즌, 시리즈'의 경우는 QT_ORDER로 태깅한다.

4.7.4. AFA_ART_CRAFT

4.7.4.1. 정의 및 설명

- 미술 작품명
 - 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

- (태깅 대상 - 화첩) 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.
- (태깅 대상 - 비, 비문) “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
 - 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

- (태깅 대상 - 앨범명) 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>
- (태깅 대상 제외 - 판소리, 종묘제례악) 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.

4.8. AF_WARES

- 판매되는 상품/제품의 명칭

4.8.1. AFW_SERVICE_PRODUCTS

4.8.1.1. 정의 및 설명

- 서비스업의 상품 명칭
 - 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, 호캬스, 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

- (태깅 대상 제외 – 보통명사) 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
 - 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 소프너, (에스티로더) 어드밴스드 나이트 리페어, (겔랑) 메테오리트 펄 일루미네이팅 파우더 등
- 향수 제품명: (샤넬) NO.5, (존 바바토스) 아티산, (다비도프) 쿨워터 맨, (불가리) 레젠크 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), 담채(김치냉장고), 애니콜 가로본능, 아이폰SE2, 갤럭시Z플립 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.
- 멀티슈명
 - 최근 '<행복한 멀티슈: 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 휴대 전화인 모델명 SCH-V500(SKT)의 다른 이름이다.

4.8.2.4. 주의 사항

- (태깅 대상 - 고유 색상명) '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_OTHER_PRODUCTS>가 요즘 나의 최애템.
 - 로지 코랄은 생얼 같은 자연스러움을 연출해 줍니다. → <로지 코랄: AFW_OTHER_PRODUCTS>은 생얼 같은 자연스러움을 연출해 줍니다.
 - 제품명은 FLOW의 '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.
- (태깅 대상 제외 -교통수단, 악기의 이름, 무기) 별도의 항목으로 분류되어 있는 무기, 교통수단, 악기의 이름은 각각 AF_TRANSPORT, AF_MUSICAL_INSTRUMENT, AF_WEAPON으로 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.
 - 무기, 교통수단, 악기에 해당하는 보통명사와 상품명 모두 AFW_OTHER_PRODUCTS로 태깅하지 않는다.
- (태깅 대상 제외 - 전자기기 보통명사) 전자제품 중 정확한 상품명으로 나온 경우만 AFW_OTHER_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 제품들은 TMI_HARDWARE로 태깅한다.
- 아이폰3GS는 한국의 스마트폰 시장을 흔들어 놓았다. → <아이폰3GS: AFW_OTHER_PRODUCTS>는 한국의 <스마트폰: TMI_HW> 시장을 흔들어 놓았다.
- (태깅 대상 제외 - 모델명) 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.
- (태깅 대상 제외 - 의류 보통명사) 의류, 스포츠용품 등 정확한 상품명에 쓰였을 경우에만 AFW_OTHER_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 것들은 CV_CLOTHING로 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.
- (태깅 단위 - 브랜드명+상품명) 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.

5. ORGANIZATION

- ORGANIZATION은 기관 및 단체 명칭을 대상으로 한다.

분류 명칭	정의 및 설명
5.1. OGG_ECONOMY	- 경제 관련 기관 및 단체 - 회사 및 기업 - 브랜드 이름
5.2. OGG_EDUCATION	- 교육 기관 및 단체 : 학교, 학원, 훈련소, 평생 교육 기관 - 학회 이름
5.3. OGG_MILITARY	- 군사 기관 및 단체 - 군사 유형 : 해군, 공군, 육군, 인민군 - 테러 조직 및 무장 단체
5.4. OGG_MEDIA	- 미디어 기관 및 단체 - 방송 관련 회사 및 기업
5.5. OGG_SPORTS	- 스포츠 기관 및 단체 - 스포츠 팀 및 구단 이름 - 스포츠 리그 - 인터넷게임 팀
5.6. OGG_ART	- 예술 기관 및 단체 - 공연장, 전시장, 박물관, 미술관, 극장
5.7. OGG_MEDICINE	- 의학/의료 기관 및 단체 - 병원, 약국
5.8. OGG_RELIGION	- 종교 기관 및 단체 - 종교 및 종파
5.9. OGG_SCIENCE	- 과학 기관 및 단체 - 연구소 및 연구원
5.10. OGG_LIBRARY	- 도서관 - 도서관 관련 기관 및 단체
5.11. OGG_LAW	- 법률 기관 및 단체
5.12. OGG_POLITICS	- 정부 및 행정 기관 - 공공기관, 정치 기관
5.13. OGG_FOOD	- 음식 관련 업체 및 회사
5.14. OGG_HOTEL	- 숙박 관련 업체 및 회사 - 호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장
5.15. OGG_OTHERS	- OGG 계열에 속하지 않는 기타 기관 및 단체

- 주의 사항
 - (태깅 대상 제외 - 보통명사) 구체적인 기관명이 아닌 보통명사는 태깅하지 않는다.
 - 도서관에서 대여한 책
 - 이때 '도서관'은 특정 기관을 지시하지 않기 때문에 태깅하지 않는다.
 - 오늘부터 모든 백화점들이 특별 할인 행사를 시작했다.
 - '백화점'은 보통명사이기 때문에 태깅하지 않는다.
 - 집 근처에는 동물병원이 없어서
 - '동물 병원'은 구체적인 병원의 이름이 없기 때문에 태깅하지 않는다.
 - (태깅 단위 - 세부 조직) 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.
 - 공정거래위원회 사무처 → <공정거래위원회: OGG_ECONOMY> 사무처
 - '사무처'는 세부 조직에 해당하기 때문에 <공정거래위원회 사무처: OGG_ECONOMY>로 태깅하지 않는다.
 - (태깅 범주 - 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>
- (태깅 범주 - OGG 계열 / AF_BUILDING / LC_OTHERS) 기관 및 단체가 OGG 계열, AF_BUILDING, LC_OTHERS로 중복 태깅이 가능할 경우 문맥과 술어를 고려하여 태깅한다. 단, OGG 계열이 나왔을 때 개별 상점의 경우 '~에 가다, ~에서 만나다' 등의 술어가 나왔다고 하더라도 OGG 계열로 태깅한다. 일정한 구역에 여러 상점이 한 곳에 모여 있는 복합문화공간, 쇼핑몰의 경우에만 한정하여 LC_OTHERS로 태깅한다. '~에 다니다, ~을 경영하다, ~이 흥하다, ~이 잘 되다, ~을 차리다'와 같은 술어와 함께 쓰이면 OGG로 태깅하고 '~에서 만나다, ~에서 구입하다'와 같은 술어와 함께 쓰일 경우 LC로 태깅한다. 또한 '~이 보이다/흔들리다, ~을 짓다/만들다'와 함께 쓰이며 건축물의 의미로 해석되는 경우 AF로 태깅한다.
 - 퇴근 후에 다이소에 갔다. → <다이소: OGG_ECONOMY>에 갔다.
 - 스타벅스에 가서 산 텀블러. → <스타벅스 : OGG_FOOD>에 가서 산 텀블러.
 - CGV에 가서 봤던 영화가 → <CGV : OGG_ART>에 가서 봤던 영화가
 - '다이소, 스타벅스, CGV'가 '-에 가다'의 서술어와 쓰여 LC로 볼 수 있으나 OGG 계열의 개별 상점이므로 OGG로 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.

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
 - 국제통화기금, 경제협력개발기구(OECD), 금융투자협회, 한국거래소, 한국신용평가
- 회사 및 기업
 - 공기업 및 사기업: 인천국제공항공사, 한국도로공사, 동양척식주식회사, 삼성전자, LG그룹, 기아차, 한화메리야스, (주)비와이씨
 - 은행, 카드사, 보험사, 증권사: 카카오뱅크, 세계은행, 한국은행, 국민은행, 비씨카드, LG 화재, AIG, HMC 투자증권
 - 개인 사업체 및 자영업: 멍스토랑, 뉴월드부동산, 열공 스타디카페, 제이엠 싱크대, 에펠(수제화)
 - 영화사, 엔터테인먼트사: CJ 엔터테인먼트, 굿프레이밍, 유니버설 스튜디오, JYP, SM 엔터테인먼트
- 브랜드 이름
 - 빈폴, 타이힐피거, 닥스, 에르메스, 구찌

5.1.3. 예문

- 경제 관련 기관 및 단체
 - <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 <한화메리야스: OGG_ECONOMY>로 출발하여 1955년 <한화산업(주): OGG_ECONOMY>으로 상호를 바꾸었고 1957년 백양상표를 등록하였다. 1960년 <한화물산(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 변경하였

다.

- 1일 오전 9시 24분 현재 유가증권시장에서 <삼성전자: OGG_ECONOMY>는 전 거래일 대비 175%(1만 3,000원) 오른 75만 7,000원에 거래 중이다.
- <휴렛팩커드: 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 우주 로케이션 장편 극 영화' 계약 논의 중.
 - <JYP엔터테인먼트: OGG_ECONOMY>가 발표하는 곡마다 1위에 오르며 핫한 여세를 보내고 있다.
- 브랜드명
 - 공방패션인데..... <샤넬: OGG_ECONOMY> 가방이랑.

5.1.4. 주의 사항

- (태깅 대상 - 운동 관련 기관) 운동과 관련된 휘트니스, 필라테스, 헬스장 등은 OGG_ECONOMY로 태깅한다. 해당 기관들은 운동 교육 기관으로 생각하여 OGG_EDUCATION으로 생각할 수 있으나 교육 이외에 개인 이용이 가능하다는 면에서 OGG_ECONOMY로 태깅한다.
 - 휴메이크 휘트니스에서 할인 행사를 시작했다. → <휴메이크 휘트니스: OGG_ECONOMY>에서 할인 행사를 시작했다.
- (태깅 대상 - 제약 회사명) 제약 회사명의 경우 OGG_ECONOMY로 태깅한다. OGG_MEDICINE으로 태깅하지 않는다.
 - 동국제약에서 출시한 이번 제품은 → <동국제약: OGG_ECONOMY>에서 출시한 이번 제품은
- (태깅 대상 제외 - 공공/정부 기관) 경제 관련 기관 중 공공 기관 및 정부 기관은 OGG_POLITICS로 태깅한다. OGG_ECONOMY로 태깅하지 않는다.
 - 근로복지공단, 한국산업단지공단 → <근로복지공단: OGG_POLITICS>, <한국산업단지공단: OGG_POLITICS>
 - '근로복지공단, 한국산업단지공단'은 공공 기관 및 정부 기관에 해당하기 때문에 OGG_ECONOMY로 태깅하지 않는다.
- (태깅 대상 제외 - 산업단지) '구로공단'과 같은 산업단지는 OGG_ECONOMY로 태깅하지 않고 LC_OTHERS로 태깅한다.

- 구로공단 → <구로공단: LC_OTHERS>
 - '구로공단'은 산업단지이기 때문에 OGG_ECONOMY로 태깅하지 않는다.
- (태깅 대상 제외 - 방위산업체) 특정한 기업을 지칭하는 것이 아니기 때문에 개체명 대상이 아니므로 태깅하지 않는다.
 - 이번 방위산업체 지원 기간에 대한 공지가 났다.
 - '방위산업체'가 특정한 기업을 나타내지 않기 때문에 OGG_ECONOMY로 태깅하지 않는다.
- (태깅 대상 제외 - 경제 회의/포럼) 'G20'과 같은 경제 모임은 OGG_ECONOMY로 태깅하나 'G20 정상 회의, 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.
- (태깅 단위 - 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>
- (태깅 단위 - 줄임말) 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, 신한 등 5대 시중 은행의 8월 말 중소기업 대출 잔액은 223조 4,364억 원으로 전달보다 1조 4,935억 원 감소했다. → <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, 국가감
 - 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 - 군사 교육 관련 기관) 군사학교나 훈련 기관과 같이 군사 교육을 시키는 기관의 명칭도 OGG_EDUCATION으로 태깅한다.
 - 웨스트포인트사관학교 → <웨스트포인트사관학교: OGG_EDUCATION>
 - 광주특수전교육단 → <광주특수전교육단: OGG_EDUCATION>
- (태깅 대상 - 훈련소) 훈련소는 모두 OGG_EDUCATION으로 태깅한다. '고용훈련소'와 같이 특정 이름이 부여되지 않은 경우라도 OGG_EDUCATION으로 태깅한다. 이때는 보통명사인 '훈련소'도 태깅한다.
 - 고용훈련소 → <고용훈련소: OGG_EDUCATION>
 - 훈련소 → <훈련소: OGG_EDUCATION>
- (태깅 대상 제외 - 학과, 단과대학) 학과나 단과대학은 학교에 포함되어 있기 때문에 OGG_EDUCATION으로 태깅하지 않는다.
 - 1954년 부산대학교 법학과를 졸업하였으며 → 1954년 <부산대학교: OGG_EDUCATION> 법학과를 졸업하였으며
 - '법학과'를 빼도 '부산대학교'라는 큰 의미는 변화가 없기 때문에 '법학과'를 포함하여 OGG_EDUCATION으로 태깅하지 않는다.
 - 홍익대 미대 → <홍익대: OGG_EDUCATION> 미대

- '미대'는 홍익대학교에 포함이 되어 있는 개념이기 때문에 OGG_EDUCATION으로 태깅하지 않는다.
- 한양공대 → <한양: OGG_EDUCATION> 공대
 - 대학명에 '-대'가 포함되지 않고 단과대가 있을 경우, 대학명만 태깅한다.
- (태깅 대상 제외 - 유네스코, 유엔) 유네스코, 유엔과 같이 교육, 문화, 사회, 정치, 과학 등의 여러 분야와 관련된 기관은 OGG_EDUCATION으로 태깅하지 않는다.
 - 유네스코에서 지속 가능한 발전을 강구하기 위한 포럼을 개최하였다. → <유네스코: OGG_OTHERS>에서 지속 가능한 발전을 강구하기 위한 포럼을 개최하였다.
 - '유네스코'는 교육뿐만 아니라 다양한 분야와 관련되어 있기 때문에 OGG_EDUCATION으로 태깅하지 않는다.
- (태깅 단위 - 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>

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 넘겼다면 <31사단: OGG_MILITARY> 헌병대와 경

찰이 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

- (태깅 대상 - 독립운동 단체) 독립운동 단체 중 전투 단체는 OGG_MILITARY로 태깅한다.
 - 오적암살단, 대한독립군 → <오적암살단: OGG_MILITARY>, <대한독립군: OGG_MILITARY>
- (태깅 대상 제외 - 군사 학교, 군사 교육 및 훈련 기관) 군사 학교와 군사 교육 기관은 OGG_MILITARY로 태깅하지 않는다.
 - 해군사관학교, 광주특수전교육단 → <해군사관학교: OGG_EDUCATION>, <광주특수전교육단: OGG_EDUCATION>
 - 해당 기관들은 '교육 기관'의 의미가 강하기 때문에 OGG_MILITARY가 아닌 OGG_EDUCATION으로 태깅한다.
- (태깅 범주 - OGG_MILITARY / AF_TRANSPORT) 군함은 문맥에 따라 AF_TRANSPORT와 OGG_MILITARY로 태깅한다. 문맥적으로 판단이 불가능할 경우 AF_TRANSPORT로 태깅한다.
 - <광개토대왕함: OGG_MILITARY>은 경기도 구리시와 자매결연을 맺고 정기적으로 지역 봉사활동에 참여하기로 했다.
 - 이때 '광개토대왕함'은 그 함선을 의미하는 것이 아니라 해당 기관을 의미하는 것이기 때문에 OGG_MILITARY로 태깅한다.
 - 한국의 구축함으로는 <광개토대왕함: AF_TRANSPORT>과 <양만춘함: AF_TRANSPORT> 등이 있다.
 - 위 문장에서 '광개토대왕함'과 '양만춘함'은 기관의 의미가 아니기 때문에 OGG_MILITARY로 태깅하지 않는다.

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, 공익광고 협의회, 월스트리트저널, 로이터통신, AP통신, 조선중앙통신, 코엑스 미디어센터
- 방송 관련 회사 및 기업
 - 스카이라이프(주)

5.4.3. 예문

- 미디어 기관 및 단체
 - 국립국어원과 <SBS: OGG_MEDIA>는 10월 12일 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 3월 부터 본격적인 위성방송 준비를 시작하여

5.4.4. 주의 사항

- (태깅 범주 - OGG_MEDIA / AFA_DOCUMENT) 신문사, 잡지사 등이 기관의 의미를 나타낼 때는 OGG_MEDIA로 태깅하고 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.
 - 위 문장에서 '조선일보'는 '신문'의 의미이기 때문에 OGG_MEDIA가 아닌 AFA_DOCUMENT로 태깅한다.

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, 맨체스터 유나이티드
- 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 재창설하고 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 프릭스: OGG_SPORTS>의 미드 라이너 '플라이' 송용준은 서머 시즌 준수한 활약을 펼쳤다.

5.5.4. 주의 사항

- (태깅 대상 - OO대표팀) 'OO대표팀'과 같은 보통명사도 OGG_SPORTS로 태깅하지만 특정 종목 없이 '대표팀, 국가대표팀'으로만 쓰였을 경우는 태깅하지 않는다.
 - 양궁 대표팀 → <양궁 대표팀: OGG_SPORTS>
 - 한국 축구 국가 대표팀 → <한국: LCP_COUNTRY> <축구 국가 대표팀: OGG_SPORTS>
 - 한국 국가 대표팀 → <한국: LCP_COUNTRY> 국가 대표팀

- 특정 종목이 쓰이지 않았기 때문에 '국가 대표팀'은 OGG_SPORTS로 태깅하지 않는다.
- (태깅 대상 - OO리그) 1부 리그 등은 구체화된 경우에만 OGG_SPORTS로 태깅한다.
 - 잉글랜드 3부 리그 → <잉글랜드 3부 리그: OGG_SPORTS>
 - 2부 리그로 하락한 팀 → 2부 리그로 하락한 팀
 - 어떤 '2부 리그'인지 확인할 수 없기 때문에 OGG_SPORTS로 태깅하지 않는다.
- (태깅 대상 제외 - 우승팀, 준우승팀) 우승팀, 준우승팀과 같은 보통명사는 태깅하지 않는다.
 - 전년도에 우승팀으로 뽑힌 팀이 시상을 한다. → 전년도에 우승팀으로 뽑힌 팀이 시상을 한다.
 - '우승팀'은 보통명사이기 때문에 OGG_SPORTS로 태깅하지 않는다.
- (태깅 범주 - OGG_SPORTS / AF_BUILDING) 골프장, 수영장 등과 같은 운동 시설이 건물이 아닌 기관의 의미로 쓰일 때에는 OGG_SPORTS로 태깅한다.
 - 여의도 한강 공원 수영장은 많은 피서객들을 유치하기 위해 시설 정비에 들어갔다. → <여의도 한강 공원 수영장: OGG_SPORTS>은 많은 피서객들을 유치하기 위해 시설 정비에 들어갔다.
 - 서강대교를 지나면 여의도 한강 공원 수영장이 보인다. → 서강대교를 지나면 <여의도 한강 공원 수영장: AF_BUILDING>이 보인다.
 - 위 문장에서 '여의도 한강 공원 수영장'은 기관이 아닌 건물의 의미이기 때문에 AF_BUILDING으로 태깅한다.
- (태깅 범주 - OGG_SPORTS / OGG_POLITICS / LCP_COUNTRY / DT 계열) 문맥상 '국가'가 스포츠 팀을 의미할 때는 OGG_SPORTS로 태깅한다. '국가'를 의미할 때는 LCP_COUNTRY, 정부를 의미할 때는 OGG_POLITICS, '왕조/시대'를 의미할 때는 DT_DYNASTY로 태깅한다.
 - 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 스포츠 팀을 의미하므로 OGG_SPORTS로 태깅한다.
 - 韓美의 주요 관계자들은 한국의 평화를 위해 노력하기로 합의했다. → <韓美 : OGG_POLITICS>의 주요 관계자들은 <한국: LCP_COUNTRY>의 평화를 위해 노력하기로 합의했다.
 - 韓美는 문맥상 정부를 의미하므로 OGG_POLITICS로 태깅하고 '한국'은 국가를 의미하기 때문에 LCP_COUNTRY로 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 DT_DYNASTY로 태깅한다.

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 범주 - OGG_ART / AF_BUILDING) 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_BUILDING로 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 평리 위안이 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 제외 – 의학 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
- 센정형외과에서 다리 수술을 했다. → <센정형외과: OGG_MEDICINE>에서 다리 수술을 했다.
 - '정형외과'와 '재활의학과'는 특정한 병원명이 아니기 때문에 OGG_MEDICINE으로 태깅하지 않는다.

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 첫 걸음으로 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

- (태깅 대상 제외 - 복합명사) 종교 기관이 복합명사의 일부이고, 종교 기관만을 독립적인 의미로 분리하기 힘든 경우에는 OGG_RELIGION으로 태깅하지 않는다.
 - 구가톨릭교도 → 구가톨릭교도
 - '구가톨릭교도'는 분리하기 힘들기 때문에 OGG_RELIGION으로 태깅하지 않는다.
- (태깅 대상 제외 - 유교) 유교는 종교 단체가 아니라 철학사상, 일종의 학문이므로 FD_HUMANITIES로 태깅한다.
 - 충과 효를 바탕으로 하는 유교 → 충과 효를 바탕으로 하는 <유교: FD_HUMANITIES>
 - '유교'는 종교 단체가 아니기 때문에 OGG_RELIGION으로 태깅하지 않는다.
- (태깅 대상 제외 - YMCA) YMCA는 청소년의 인격과 지도력 훈련 교육 및 청소년 문화사업을 중점으로 하는 기관으로 OGG_OTHERS로 태깅한다.
 - YMCA에서 주관하는 이 행사 → <YMCA: OGG_OTHERS>에서 주관하는 이 행사
 - 'YMCA'는 종교 단체가 아니기 때문에 OGG_RELIGION으로 태깅하지 않는다.

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, 독도수산연구센터

5.9.3. 예문

- 과학 기관 및 단체
 - 1999년 법정법인인 <정보통신연구진흥원: OGG_SCIENCE>으로 새로 출범한 뒤 그해 해외 정보통신 기술협력센터와 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.

5.10. 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_LIBRARY>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 단위 - 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 전무가 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 공무상요양불승인처

분 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

- (태깅 대상 - 법원) 법원은 OGG_POLITICS와 OGG_LAW에 모두 해당되는데 OGG_LAW로 태깅한다.
 - 대법원 → <대법원: OGG_LAW>
- (태깅 단위 - 국가명+지명+법원) 국가명과 지명, 법률 기관이 함께 쓰인 경우, 국가명은 LCP_COUNTRY로 따로 태깅하고 지명과 법원은 OGG_LAW로 함께 태깅한다.
 - 대한제국 법무 전라북도재판소 → <대한제국: LCP_COUNTRY> <법부: OGG_LAW> <전라북도재판소: OGG_LAW>
- (태깅 단위 - 국가 기관+법률 기관) 국가 기관과 법률 기관이 함께 쓰인 경우, 국가 기관은 OGG_POLITICS로 태깅하고 법률 기관은 OGG_LAW로 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
- 공공기관
- 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
- 공공기관
 - 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 <예산처: 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: OGG_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 정책질의를 착수했습니다.

5.12.4. 주의 사항

- (태깅 대상 - 독립운동 기관) 독립운동 기관도 일종의 정치 기관으로 간주하여 OGG_POLITICS로 태깅한다.
 - 간도 국민회, 간도대한민회 → <간도 국민회: OGG_POLITICS>, <간도대한민회: OGG_POLITICS>
 - 단, 독립운동 단체 중에 전투 단체는 OGG_MILITARY로 태깅한다.
- (태깅 대상 - 국가 연구기관) 국가 연구 기관도 OGG_POLITICS로 태깅한다.
 - 집현전 → <집현전: OGG_POLITICS>
- (태깅 대상 - 조선시대 정치 집단) 서론, 노론 등 조선시대 정치 집단도 OGG_POLITICS로 태깅한다.
 - 서인에서 갈라져 나온 노론 → 서인에서 갈라져 나온 <노론: OGG_POLITICS>
- (태깅 대상 - 여야 통합 정부) '여야 통합 정부', '여야 정부'는 전체를 OGG_POLITICS로 태깅한다.
 - 이번 대선 주자는 여야 통합 정부를 만들겠다고 했다. → 이번 대선 주자는 <여야 통합 정부: OGG_POLITICS>를 만들겠다고 했다.
- (태깅 대상 - 줄임말) '시, 군, 구'가 줄임 표현으로 나올 시 기관의 의미를 나타내면 OGG_POLITICS로 태깅한다.
 - 양구군은 (중략) 군에서 → <양구군: OGG_POLITICS>은 (중략) <군: OGG_POLITICS>에서
 - '시, 군, 구'가 단독으로 나와도 기관의 의미를 가질 경우에는 OGG_POLITICS로 태깅한다.
- (태깅 대상 제외 - 포괄적인 단어) 기관과 연관된 단어가 아니며, 특정한 의미를 가지지 않는 포괄적인 단어인 경우에는 태깅 대상이 아니다.
 - 임시회, 정기회, 합의회행정기관, 하급행정기관
- (태깅 대상 제외 - 국가 과학 기관) 국가 연구기관 중에 과학 기관에 해당하는 것은 OGG_POLITICS로 태깅하지 않고 OGG_SCIENCE로 태깅한다.
 - 국립수산물과학원 → <국립수산물과학원: OGG_SCIENCE>
- (태깅 단위 - 국가+정부/국가+도시+정부 기관/사람 이름+정부) '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>
- (태깅 단위 - LC 계열 + OG 계열) LC 계열이 단순히 기관이 소재한 장소를 의미할 경우에는 각각 태깅한다. 그러나 LC 계열을 포함하여 하나의 기관 이름으로 기능할 경우에는 전체를 함께 OGG_POLITICS로 태깅한다.
 - 미국 컬럼비아대학 → <미국: LCP_COUNTRY> <컬럼비아대학: OGG_EDUCATION>
 - 컬럼비아 대학이 미국에 위치한다는 뜻이므로 나누어 태깅한다.
 - 남구을 지구당 위원장 → <남구을: 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 OGG_POLITICS로 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.
- (태깅 범주 - OGG_POLITICS / CV_OCCUPATION) 경찰은 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.
 - 경찰에 신고하다. → <경찰: OGG_POLITICS>에 신고하다.
 - '교통'이라고 적힌 연두색 조끼를 입은 경찰을 배치했다. → '교통'이라고 적힌 연두색 조끼를 입은 <경찰: CV_OCCUPATION>을 배치했다.
 - 위 문장에서 '경찰'은 경찰 공무원을 의미하기 때문에 OGG_POLITICS로 태깅하지 않고 CV_OCCUPATION으로 태깅한다.
- (태깅 범주 - OGG_POLITICS / LC 계열) 맥락에 따라 OGG_POLITICS가 될 수 있고 LC가 될 수도 있고 LC 계열이 될 수도 있는데, 그 판단 기준은 해당 문장의 서술어를 중심으로 문장 의미에 따라 결정한다. '~사람, ~인'과 같은 명사나 '~으로 귀화하다, ~으로 망명하다'와 같은 술어와 같이 쓰일 경우에는 LCP_COUNTRY 계열로 태깅한다.
 - <프랑스: OGG_POLITICS>도 '방송장악' 논란
 - <러시아: LCP_COUNTRY>, <프랑스: LCP_COUNTRY> 같은 쟁쟁한 잠수함 강국들은 물론 '스승'이라 할 <독일: LCP_COUNTRY>까지 물리치고 거둔 성과여서 더욱 대견스럽다.
 - 위 문장에 나타난 나라명은 각 나라의 정부 기관을 의미하지 않기 때문에 OGG_POLITICS로 태깅하지 않고 LC 계열로 태깅한다.
 - 한편 <고양시: OGG_POLITICS>는 내년 10월에 열리는 제92회 전국체전도 유치했다.
 - <고양시: LCP_CITY> 덕양구 행신동 서정초등학교 인근 아파트형 공장과 관련해
 - 위 문장에서 '고양시'는 행정구역으로 쓰였기 때문에 OGG_POLITICS로 태깅하지 않고 LC계열로 태깅한다.
- (태깅 범주 - OGG_POLITICS / OGG_SPORTS / LC 계열 / DT 계열) 문맥상 정부를 의미할 때는

OGG_POLITICS, 스포츠 팀을 의미할 때는 OGG_SPORTS, '국가'를 의미할 때만 LCP_COUNTRY, '왕조/시대'를 의미할 때는 DT_DYNASTY로 태깅한다.

- 韓美의 주요 관계자들은 한국의 평화를 위해 노력하기로 합의했다. → <韓美: OGG_POLITICS>의 주요 관계자들은 <한국: LCP_COUNTRY>의 평화를 위해 노력하기로 합의했다.
 - '한국'은 문맥상 국가를 의미하므로 LCP_COUNTRY로 태깅한다.
- 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 스포츠 팀을 의미하므로 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 DT_DYNASTY로 태깅한다.

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, 도미노피자

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 컵 '비아(Via)' 출시를 통해 국내 인스턴트 커피 시장에 본격 진출한다.
 - 우선 저희 집은 경기도 포천에서 작은 <바위식당: OGG_FOOD>이라는 식당을 운영하고 있습니다.

5.13.4. 주의 사항

- (태깅 대상 - 음식 관련 업체 및 회사) 본죽, 맥도날드, 스타벅스 등이 OGG_ECONOMY로 여겨지기도 하지만 모든 음식과 관련된 업체 및 회사는 모두 OGG_FOOD로 태깅한다.
- (태깅 대상 - 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.
- (태깅 단위 - 상호명+OO점) '스타벅스 홍대점'과 같이 '상호명 + OO점'이 나올 경우 하나로 묶어서 태깅한다.
- 스타벅스 홍대점 → <스타벅스 홍대점: OGG_FOOD>

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

- (태깅 대상 - 산장) 산장의 경우 구체적인 상호명과 함께 쓰였을 경우 태깅한다.
 - 설악산장은 사람이 많을 수도 있으니 → <설악산장: OGG_HOTEL>은 사람이 많을 수도 있으니
 - 속초 산장이 몇 군데가 있는지 확인해보고 → <속초: LCP_CITY> 산장이 몇 군데가 있는지 확인해보고
 - '속초 산장'은 상호명이 아닌 '속초'에 위치한 '산장'이라는 의미이기 때문에 OGG_HOTEL로 태깅하지 않는다.

- (태깅 대상 제외 - 룸 이름) 호텔 내 존재하는 룸 이름은 태깅하지 않는다.
 - 롯데호텔 주니어 스위트룸 → <롯데호텔: OGG_HOTEL> 주니어 스위트룸
 - '주니어 스위트 룸'은 '롯데호텔'에 속한 것이기 때문에 OGG_HOTEL로 태깅하지 않는다.
- (태깅 범주 - OGG_HOTEL / LC_OTHERS) 휴양시설은 OGG_HOTEL과 LC_OTHERS가 모두 가능한데 '호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장'의 특정 명칭만 OGG_HOTEL로 태깅하고 그 외 휴양시설인 '야영장, 체험장'등의 명칭은 LC_OTHERS로 태깅한다.
 - 이번 휴가는 <신라호텔: OGG_HOTEL>로 가기로 했다.
 - <용연 청소년 야영장: LC_OTHERS>은 청소년들에게 건전한 야영 활동과 심신 수련 및 호연지기를 키워 주기 위하여 건립했다.

5.15. OGG_OTHERS

5.15.1. 정의 및 설명

- OGG 계열에 속하지 않는 나머지 모든 기관 및 단체

5.15.2. 분석 대상

- 유색인종향상전국협회, 국제소비자연맹
- 비세그라드 그룹
- 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.
- <비세그라드 그룹: OGG_OTHERS>은 빠른 경제 성장을 보였다.
- 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 바

람직하지 않다면서도 청와대와 국회의 확실한 분권과 지방 자치 실현을 위해서는 제도를 법률적으로 바꾸면 된다고 말했습니다.

5.15.4. 주의 사항

- (태깅 대상 - UN, 유네스코) 교육, 문화, 정치, 사회, 과학 등의 여러 분야와 관련된 기관으로 OGG_OTHERS로 태깅한다.
 - 유네스코에서 지속 가능한 발전을 강구하기 위한 포럼을 개최하였다. → <유네스코: OGG_OTHERS>에서 지속 가능한 발전을 강구하기 위한 포럼을 개최하였다.
- (태깅 대상 - 국경수비대) '국경수비대'는 나라에 따라 군사 조직인 경우와 경찰 조직인 경우로 나뉘기 때문에 OGG_OTHERS로 태깅한다.
 - 국경수비대에 잡혀 이송되고 있다. → <국경수비대: OGG_OTHERS>에 잡혀 이송되고 있다.

6. LOCATION

- LOCATION은 지역/장소와 지형/지리 명칭 등을 대상으로 한다. 지역/ 장소는 공식적 지역 명칭과 행정 구역 명칭을 나타내며 LCP(LoCation Place) 계열로 표시한다. 지형/ 지리 명칭은 국가명이나 행정구역 명과는 별개인 산, 강과 같은 자연에 대한 지리 명칭을 나타내며, LCG(LoCation Geography) 계열로 표시한다. 그 외 천체 명칭은 LC_SPACE, LC 계열의 세부 유형이 아닌 기타 장소는 LC_OTHERS로 태깅한다.

분류 명칭	정의 및 설명
6.1. LCP_COUNTRY	- 국가명 및 왕국의 명칭 - 고대 국가명
6.2. LCP_PROVINCE	- 도, 도에 해당하는 외국의 지역명(주, 성) - 특별자치구역이나 대도시권 - 지방/지역명
6.3. LCP_COUNTY	- 군, 구, 동, 면, 읍, 리 등과 같은 세부 행정구역명 - 이에 해당하는 외국의 지역명
6.4. LCP_CITY	- 국가의 수도를 제외한 도시명
6.5. LCP_CAPITALCITY	- 수도명
6.6. LCG_RIVER	- 강, 호수/육수의 이름
6.7. LCG_OCEAN	- 해양, 바다, 해역의 명칭
6.8. LCG_BAY	- 반도, 곶, 만, 피오르의 명칭
6.9. LCG_MOUNTAIN	- 산/산맥, 능선, 광산, 봉(우리)/령, 고개/재의 명칭
6.10. LCG_ISLAND	- 섬, 제도, 환초/산호섬의 명칭
6.11. LCG_CONTINENT	- 대륙 및 대륙의 일부분의 명칭
6.12. LC_SPACE	- 항성, 행성의 명칭 - 위성/유성 및 혜성의 명칭 - 별자리, 성운, 달의 명칭
6.13. LC_OTHERS	- LC 계열의 세부 유형에 속하지 않는 기타 장소의 명칭

- 주의 사항
 - (태깅 대상 - 구체적인 지형 명칭) 지형 명칭에 지형의 유형은 포함되지 않고, 구체적인 지형 명칭만 해당된다.
 - <오세아니아: LCG_CONTINENT>는 세계에서 가장 작은 대륙입니다.
 - 구체적인 지형 명칭인 '오세아니아'만 태깅하고 지형의 유형 명칭인 '대륙'은 태깅하지 않는다.
 - (태깅 대상 - 정식 명칭) 건축물이 지명과 함께 쓰인 경우, 전체가 정식 명칭이 아니면 분리하여 지

명은 LC로 태깅하고, 정식 명칭이면 분리하지 않고 하나로 AF_BUILDING으로 태깅한다.

- 한강 잠원동 수영장 → <한강: LCG_RIVER> <잠원동: LCP_COUNTY> 수영장

- '한강 잠원동 수영장'은 정식 명칭이 아니므로 분리하여 태깅한다. '수영장'은 태깅 대상이 아니므로 태깅하지 않는다.

- 잠원 한강공원 수영장 → <잠원 한강공원 수영장: AF_BUILDING>

- '잠원 한강공원 수영장'이 정식 명칭이므로 분리하지 않고 전체를 AF_BUILDING으로 태깅한다.

- (태깅 대상 제외 - 가상의 장소) 속담이나 게임, 가상의 지역/ 장소, 지형/ 지리 명칭은 태깅하지 않는다.

- (태깅 단위 - 주소) 지역/장소명은 세부적으로 분리하여 태깅한다.

- 충청남도 예산군 덕산면 상가리 37 → <충청남도: LCP_PROVINCE> <예산군: LCP_COUNTY> <덕산면: LCP_COUNTY> <상가리: LCP_COUNTY> <37: QT_ADDRESS>

- (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>

- 서울시 강남구 삼성로 1길 150 대치미도 1차 아파트 103-101 → 서울시 강남구 삼성로 1길 150 <대치미도 1차 아파트: LC_OTHERS> <103-101: QT_ADDRESS>

- 공동주택의 동/호수를 'n-n'의 형식으로 묶어 쓴 경우에는 위와 같이 QT_ADDRESS로 태깅한다.

- (태깅 단위 - 지명+법률 기관) 국가명과 도시명, 법률 기관이 함께 쓰인 경우, 국가명은 LCP_COUNTRY로 따로 태깅하고 도시명과 법원은 OGG_LAW로 함께 태깅한다.

- 대한제국 법부 전라북도재판소 → <대한제국: LCP_COUNTRY> <법부: OGG_LAW> <전라북도재판소: OGG_LAW>

- (태깅 단위 - 국가+정부/국가+도시+정부 기관/사람 이름+정부) '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>

- (태깅 단위 - LC계열 + OG계열) LC계열이 단순히 기관이 소재한 장소를 의미할 경우에는 각각 태깅한다. 그러나 LC계열을 포함하여 하나의 기관 이름으로 기능할 경우에는 전체를 함께 OGG_POLITICS로 태깅한다.

- 미국 컬럼비아대학 → <미국: LCP_COUNTRY> <컬럼비아대학: OGG_EDUCATION>

- 컬럼비아 대학이 미국에 위치한다는 뜻이므로 나누어 태깅한다.

- 남구을 지구당 위원장 → <남구을: 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 OGG_POLITICS로 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.
- (태깅 범주 - LC계열 / OGG_POLITICS) 맥락에 따라 LC가 될 수도 있고 OG가 될 수도 있다. 기관으로 쓰였을 때만 OG로 태깅한다.
 - <프랑스: OGG_POLITICS>도 '방송장악' 논란
 - <러시아: LCP_COUNTRY>, <프랑스: LCP_COUNTRY> 같은 쟁쟁한 잠수함 강국들은 물론 '스승'이라 할 <독일: LCP_COUNTRY>까지 물리치고 거둔 성과여서 더욱 대견스럽다.
 - 위 문장에 나타난 나라명은 각 나라의 정부 기관을 의미하지 않기 때문에 OGG_POLITICS로 태깅하지 않고 LC계열로 태깅한다.
 - 한편 <고양시: OGG_POLITICS>는 내년 10월에 열리는 제92회 전국체전도 유치했다.
 - '고양시'는 전국체전을 유치한 기관으로 쓰였으므로 OGG_POLITICS로 태깅한다.
 - <고양시: LCP_CITY> 덕양구 행신동 서정초등학교 인근 아파트형 공장과 관련해
 - 위 문장에서 '고양시'는 행정구역으로 쓰였기 때문에 OGG_POLITICS로 태깅하지 않고 LC계열로 태깅한다.

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
- 고대 국가명
 - 고구려, 백제, 신라, 고려, 조선, 명나라, 원나라 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.
- 고대 국가명
 - <명나라: LCP_COUNTRY>로 사신을 보냈다.

6.1.4. 주의 사항

- (태깅 대상 - 약칭) 국가를 약칭(略稱)으로 표현한 경우도 태깅한다.
 - 韓美 국민들의 행복지수 비교 → <韓美: LCP_COUNTRY> 국민들의 행복지수 비교
- (태깅 대상 - 민주공화국) 민주공화국은 특정한 나라가 앞에 붙어 있는 경우만 개체명 대상으로 인식한다. 구체적인 국가명이 아닌 '국가/ 나라/민주공화국' 등의 보통명사는 태깅하지 않는다.
 - 자이르는 콩고민주공화국의 전 이름이다. → <자이르: LCP_COUNTRY>는 <콩고민주공화국: LCP_COUNTRY>의 전 이름이다.
 - 콩고는 지금의 민주공화국이 되기 위해 험난한 일들을 겪었다. → <콩고: LCP_COUNTRY>는 지금의 민주공화국이 되기 위해 험난한 일들을 겪었다.
 - '콩고'는 국가명이므로 LCP_COUNTRY로 태깅하고 보통명사로 쓰인 '민주공화국'은 태깅하지 않는다.
- (태깅 대상 제외 - 한반도) '한반도'에는 남한과 북한이 모두 있으므로 국가를 나타내는 LCP_COUNTRY로 태깅하지 않는다. 문맥상 지형명으로 사용된 경우만 LCG_BAY로 태깅한다. 지형명이 아닌 경우는 태깅하지 않는다.
 - 비행기가 한반도 상공을 지나고 있다. → 비행기가 <한반도: LCG_BAY> 상공을 지나고 있다.
- (태깅 단위 - 한중일, 한미) 우리말샘에 '한중일, 한미'가 한 단어로 등재되어 있으므로 '한중일, 한미' 전체를 LCP_COUNTRY로 태깅한다.
 - <한중일: LCP_COUNTRY>, <한미: LCP_COUNTRY>
- (태깅 범주 - LC_COUNTRY / OGG계열 / DT계열) 문맥상 '국가'를 의미할 때만 LCP_COUNTRY로 태깅한다. 문맥상 정부를 의미할 때는 OGG_POLITICS, 스포츠 팀을 의미할 때는 OGG_SPORTS로 태깅한다. '왕조/시대'를 의미할 때는 DT_DYNASTY로 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. → <韓美: OGG_POLITICS>의 주요 관계자들은 <한국: LCP_COUNTRY>의 평화를 위해 노력하기로 합의했다.
 - 韓美는 문맥상 정부를 의미하므로 OGG_POLITICS로 태깅한다.
 - 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 스포츠 팀을 의미하므로 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 DT_DYNASTY로 태깅한다.
- (태깅 범주 - LCP_COUNTRY / CV_LANGUAGE) '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.
 - 단순히 한국어로 된 드라마를 지칭하는 것이 아닌 한국에서 제작된 드라마를 지칭하므로 CV_LANGUAGE가 아닌 LCP_COUNTRY로 태깅한다.
 - <영어: CV_LANGUAGE> 버전으로 먼저 출간된 이 한국어 교재는 현재 <중국: CV_LANGUAGE> 버전과 <일본: CV_LANGUAGE> 버전의 번역 작업이 진행 중이다.
 - '일본'이 확실히 특정 언어들을 지칭하므로 CV_LANGUGAE로 태깅한다.

6.2. LCP_PROVINCE

6.2.1. 정의 및 설명

- 도, 도에 해당하는 외국의 지역명, 특별자치구역이나 대도시권, 지방/지역명

6.2.2. 분석 대상

- 도
 - 경기도, 강원도, 충청남도 등
- 도에 해당하는 외국의 지역명(주, 성)
 - 캘리포니아주, 애리조나주, 퀘벡주, 슈비츠주, 네바다주, 길림성, 도쿄도, 홋카이도, 오사카부, 교토부 등
- 특별자치구역이나 대도시권
 - 홍콩, 할리우드 등
- 지방/지역명
 - 호남, 영남 등

6.2.3. 예문

- 도
 - 전국의 행정구역을 13도로 구획하면서 <충청남도: LCP_PROVINCE>와 <충청북도: LCP_PROVINCE>로 나누어, 대체로 현재와 같은 행정구역의 윤곽을 지니게 되었다.
 - 이번 태풍으로 <경남: LCP_PROVINCE>과 <경북: LCP_PROVINCE>은 큰 피해를 입었다.
- 도에 해당하는 외국의 지역명(주, 성)
 - 이 <퀘벡주: LCP_PROVINCE>가 캐나다 연방에서 이탈하여 독립하려는 운동이 1960년대부터 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.
- <길림성: LCP_PROVINCE>은 중국 북동부에 위치하고 있으며 <지린성: LCP_PROVINCE>이라고도 한다.
- <오사카부: LCP_PROVINCE> 전체 인구는 약 880만 명이다.
- 특별자치구역이나 대도시권
 - 빅토리아 시대를 <홍콩: LCP_PROVINCE>의 수도로 알고 있는 사람이 있는데, 정확하게는 <홍콩: LCP_PROVINCE>의 행정적 중심지 역할을 하는 지역이다.
- 지방/지역명
 - <호남: LCP_PROVINCE> 지방에 내린 폭우

6.2.4. 주의 사항

- (태깅 대상 - 일본의 행정구역명 '도, 도, 부') 일본에서 도시보다 큰 단위로는 '도(도쿄도), 도(홋카이도), 부(오사카부, 교토부), 현(나머지 43개의 현)'이 있다. 이중 '도(도쿄도), 도(홋카이도), 부(오사카부, 교토부)'만 LCP_PROVINCE로 태깅한다. 'OO현'은 우리나라의 '도'보다 작은 개념의 행정구역으로 보고 LCP_COUNTY로 태깅한다.
 - 도쿄도, 홋카이도, 오사카부, 교토부
 - 모두 LCP_PROVINCE로 태깅한다.
 - 일본의 나가사키현은 유명한 음식이 많다. → 일본의 <나가사키현: LCP_COUNTY>은 유명한 음식이 많다.
 - 일본의 '현'은 LCP_COUNTY로 태깅한다.
- (태깅 대상 제외 - O부 지역/O부권) '호남/영남'과 같이 특정 지역명이 아닌 '중부/남부 지역'과 같이 지역의 부분을 나타내는 경우 LCP_PROVINCE로 태깅하지 않는다. '중부/남부 지역'의 경우 '중부/남부'만 TM_DIRECTION으로 태깅한다. '중부권/남부권'과 같이 'O부권'으로 사용된 경우 복합어이기 때문에 TM_DIRECTION으로 내부 태깅하지 않는다.
 - 중부 지방에 내린 폭우 → <중부: TM_DIRECTION> 지방에 내린 폭우
 - '중부'는 LCP_PROVINCE가 아닌 TM_DIRECTION으로 태깅한다.
 - 고인돌이 한반도 남부권에서 발견되었다. → 고인돌이 한반도 남부권에서 발견되었다.
 - '남부권'은 복합어이므로 태깅하지 않는다.
- (태깅 단위 - OO 지방/지역) 뒤에 '지방/지역'이 붙으면 '지방/지역'을 제외하고 태깅한다.
 - 호남지방에 내린 폭우 → <호남: LCP_PROVINCE> 지방에 내린 폭우
- (태깅 범주 - LCP_PROVINCE / LCG_ISLAND) '제주도, 홋카이도, 오키나와, 코스라에주'와 같이 섬이면서 행정구역일 경우 LCP_PROVINCE로 우선 태깅한다. 문맥상 확실히 섬일 경우만 LCG_ISLAND로 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.
 - 코스라에주는 현재 전통 농업과 어업이 주를 이룬다. → <코스라에주: LCP_PROVINCE>는 현재 전통

농업과 어업이 주를 이룬다.

6.3. LCP_COUNTY

6.3.1. 정의 및 설명

- 군, 구, 동, 면, 읍, 리 등과 같은 세부 행정구역명 및 이에 해당하는 외국의 지역명

6.3.2. 분석 대상

- 군, 구, 동
 - 영월군, 완도군, 종로구, 강서구, 노원구, 노암동, 방화동, 서천동, 종로1가동, 회현동1가, 충무로2가 등
- 면, 읍, 리
 - 도암면, 춘포면, 담양읍, 양양읍, 횡계리, 영남리 등
- 군, 구, 동, 면, 읍, 리에 해당하는 외국의 지역명
 - 미나토구, 브루클린, 나가사키현 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.
 - 부산광역시 서구의 행정동 <부민동: LCP_COUNTY>은 법정동 <부민동1가: LCP_COUNTY>, <부민동2가: LCP_COUNTY>, <부민동3가: LCP_COUNTY>, <부용동1가: LCP_COUNTY>, <부용동2가: LCP_COUNTY>를 관할한다.
- 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 행정구역 개편 때 면이 폐지되면서 통합되어 신설한 <은천군: LCP_COUNTY> <영남리: LCP_COUNTY>가 되었다가 1974년 5월 남포직할시에 이관되었다.
- 군, 구, 동, 면, 읍, 리에 해당하는 외국 지역명
 - 이나바 감독은 12일 일본 도쿄도(東京都) <미나토구: LCP_COUNTY>(<港区: LCP_COUNTY>)에 있는 그랜드프린스호텔에서

6.3.4. 주의 사항

- (태깅 대상 - 일본의 행정구역명 '현') 일본에서 도시보다 큰 단위로는 '도(도쿄도), 도(홋카이도), 부(오사카부, 교토부), 현(나머지 43개의 현)'이 있다. 이중 '현'만 우리나라의 '도'보다 작은 개념의 행정구역

으로 보고 LCP_COUNTY로 태깅한다.

- 일본의 나가사키현은 유명한 음식이 많다. → 일본의 <나가사키현: LCP_COUNTY>은 유명한 음식이 많다.
 - 일본의 '현'은 LCP_COUNTY로 태깅한다.
- 도쿄도, 홋카이도, 오사카부, 교토부
 - 모두 LCP_PROVINCE로 태깅한다.
- (태깅 대상 제외 - 도로명) 2014년부터 정식 시행된 새로운 주소 표기법인 도로명주소법에는 도로명이 주소로 표시된다. 도로명은 도시보다 작은 단위지만 LC_COUNTY로 태깅하지 않고 AF_ROAD로 태깅한다.
- 서울특별시 강서구 금남화로 154 → <서울특별시: LCP_CAPITALCITY> <강서구: LCP_COUNTY> <금남화로: AF_ROAD> <154: QT_ADDRESS>
- (태깅 대상 제외 - OO구갑/OO구을) 행정구역상 정식 명칭일 경우만 LCP_COUNTY로 태깅한다. 선거구를 나타내는 'OO구을/OO구갑' 등의 경우 행정구역상 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>, <노원병: LC_OTHERS> 등 모두 세 명의 지역구 의원이 있다.
 - '노원구'는 구의 이름이므로 LCP_COUNTY로 태깅하고, '노원갑, 노원을, 노원병'은 선거구를 나타내므로 LC_OTHERS로 태깅한다.
- (태깅 단위 - OO동/OO로 1가) '종로1가, 종로2가'의 경우 전체가 행정동의 이름이고, '회현동1가, 태평로1가, 을지로1가'의 경우 전체가 법정동의 이름이다. '1가, 2가' 등이 있어도 전체가 행정동이나 법정동의 이름이면 LC와 QT를 분리하지 않고 전체를 묶어 태깅한다.
- 충무로3가 일대에는 편집디자인 기획사들과 출력소들이 밀집되어 있다. → <충무로3가: LCP_COUNTY> 일대에는 편집디자인 기획사들과 출력소들이 밀집되어 있다.
 - '충무로3가'는 전체가 법정동 이름이므로 전체를 LCP_COUNTY로 태깅한다.
- (태깅 범주 - LCP_COUNTY / LC_OTHERS) '신촌, 강남'과 같이 도시 안의 일부 지역이면서 행정구역명으로도 쓰이는 지역은 LCP_COUNTY로 태깅한다. '신촌역, 강남역'과 같이 문맥상 '역'이 확실할 경우만 LC_OTHERS로 태깅한다. 문맥으로 판단이 어려운 경우 LCP_COUNTY로 우선 태깅한다. 한편, '서면, 광안리, 압구정로데오'의 경우 지하철역 이름으로 그 일대를 일컫기는 하지만 정확한 행정구역 명칭이 아니므로 LCP_COUNTY로 태깅하지 않고 LC_OTHERS로 태깅한다.
- 지난 주말엔 친구들과 신촌에서 모였어요! → 지난 주말엔 친구들과 <신촌: LCP_COUNTY>에서 모였어요!
- 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.
 - 도시 안의 일부 지역을 나타내므로 LCP_COUNTY로 태깅한다.
- 신촌 다음 역에서 내리면 된다. → <신촌: LC_OTHERS> 다음 역에서 내리면 된다.
 - 맥락상 '신촌역'을 나타내므로 LC_OTHERS로 태깅한다.
- 지난 주말에는 대학 동기들과 시간을 맞춰 <서면: LC_OTHERS> 맛집에 다녀왔는데요.

- (태깅 범주 - 행정구역명이 바뀐 지역) '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.
- (태깅 범주 - 동/서/남/북+도시명) '동대구/서대구'와 같이 도시의 일부를 나타내는 '동/서/남/북+도시명'인 경우 도시보다 작은 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.
- (태깅 범주 - 해운대) '해운대'와 같은 경우 해수욕장/해변, 행정구역, 대로, 지하철역을 모두 나타낼 수 있다. 주소로 사용되어 '해운대구'를 의미하면 LCP_COUNTY로 태깅한다. 해수욕장/해변일 경우 LC_OTHERS로 태깅하고, '해운대로'면 AF_ROAD, '해운대역'을 의미하면 LC_OTHERS로 태깅한다. 문맥으로 판단하기 어려운 경우 LC_OTHERS로 우선 태깅한다.
 - 부산 해운대 우동의 한 아파트에서 불이 나 주민 수십 명이 대피하고 1명이 다쳤다. → 부산 <해운대: LCP_COUNTY> 우동의 한 아파트에서 불이 나 주민 수십 명이 대피하고 1명이 다쳤다.
 - 문맥상 '해운대'가 부산의 해운대구를 뜻하므로 LCP_COUNTY로 태깅한다.
 - 부산 해운대 가 볼 만한 곳 10 → 부산 <해운대: LC_OTHERS> 가 볼 만한 곳 10
 - 문맥상 해수욕장/해변을 뜻하므로 LC_OTHERS로 태깅한다.

6.4. LCP_CITY

6.4.1. 정의 및 설명

- 국가의 수도를 제외한 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

- (태깅 대상 제외 - 수도) 국가의 수도도 도시에 해당되지만 수도는 LCP_CAPITALCITY로 우선 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.
- (태깅 대상 제외 - 동/서/남/북+도시명) '동대구/서대구'와 같이 도시의 일부를 나타내는 '동/서/남/북+도시명'인 경우 도시보다 작은 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.
- (태깅 범주 - 과거 수도) 예전에 수도였더라도 현재 없어졌거나 일반 도시면 LCP_CITY로 태깅한다. 현재 수도인 도시만 LCP_CAPITALCITY로 태깅한다. 현재 수도가 아니라도 문맥상 확실히 수도를 나타내는 경우는 LCP_CAPITALCITY로 태깅한다. 문맥으로 판단이 어려운 경우 LCP_CITY로 우선 태깅한다.
 - 아르마비르, 음반자궁구, 경주 → <아르마비르: LCP_CITY>, <음반자궁구: LCP_CITY>, <경주: LCP_CITY>
 - '아르마비르, 음반자궁구, 경주' 모두 현재 수도가 아니며 문맥상으로도 수도로 쓰이지 않았으므로 LCP_CITY로 태깅한다.
 - 당시 신라의 수도인 <경주: LCP_CAPITALCITY>에서 대규모 전투가 벌어졌다.
 - '경주'는 문맥상 확실히 신라의 수도를 나타내므로 LCP_CAPITALCITY로 태깅한다.

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.
 - '수도에의 집중현상'은 프랑스의 특징이며, <파리: LCP_CAPITALCITY>는 세계 제4위의 인구밀집 지역이다.
 - <브라질리아: LCP_CAPITALCITY>를 21세기의 도시로 건설하였다.

6.5.4. 주의 사항

- (태깅 범주 - 과거 수도) 현재 수도인 도시만 LCP_CAPITALCITY로 태깅한다. 예전에 수도였더라도 현재 없어졌거나 일반 도시면 LCP_CITY로 태깅한다. 현재 수도가 아니라도 문맥상 확실히 수도를 나타내는 경우는 LCP_CAPITALCITY로 태깅한다. 문맥으로 판단이 어려운 경우 LCP_CITY로 우선 태깅한다.
- 당시 신라의 수도인 <경주: LCP_CAPITALCITY>에서 대규모 전투가 벌어졌다.
 - '경주'는 문맥상 확실히 신라의 수도를 나타내므로 LCP_CAPITALCITY로 태깅한다.
- 아르마비르, 음반자궁구, 경주 → <아르마비르: LCP_CITY>, <음반자궁구: LCP_CITY>, <경주: LCP_CITY>
 - '아르마비르, 음반자궁구, 경주' 모두 현재 수도가 아니며 문맥상으로도 수도로 쓰이지 않았으므로 LCP_CITY로 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 감입곡류하며 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

- (태깅 대상 제외 - 4대강) 한강, 섬진강 등 특정 강의 명칭 없이 단순히 '4대강'만 나왔을 경우 태깅하지 않는다. '4대강'의 경우 '4대강 사업 국가 프로젝트'의 의미를 지닐 경우 TMI_PROJECT로 태깅한다. 줄임말로 사용된 경우라도 국가 프로젝트의 의미라면 TMI_PROJECT로 태깅한다.
 - 4대강이 녹조로 뒤덮였다. → 4대강이 녹조로 뒤덮였다.
 - '4대강'이 특정한 강을 칭하지 않으므로 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.
 - '4대강'이 국가 프로젝트를 뜻하므로 TMI_PROJECT로 태깅한다.
- (태깅 대상 제외 - OO보) '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/바다/해역의 명칭

6.7.2. 분석 대상

- 해양
 - 태평양, 인도양 등
- 바다
 - 동해, 남극해, 북극해, 데이비스해, 아문센해, 산호해 등
- 해역
 - 바람내리포, 바람내리개, 아덴만 해역 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 <토러스 해협: LCG_OCEAN>을 거쳐 <아라푸라해: LCG_OCEAN>(<海>)로 통한다.
- 해역
 - <바람내리포: LCG_OCEAN>는 지역주민들이 바람이 불 때 비탈에 바람이 부딪혀 내려와서 "<바람내리개: LCG_OCEAN>"라고 부르고 있다.
 - 부산작전기지에서 소말리아 <아덴만 해역: LCG_OCEAN>으로 출항하였습니다.

6.8. LCG_BAY

6.8.1. 정의 및 설명

- 반도, 곶, 만, 피오르의 명칭

6.8.2. 분석 대상

- 반도
 - 랴오둥반도, 변산반도, 발칸반도, 이베리아반도 등
- 곶
 - 간절곶, 호미곶 등
- 만
 - 도쿄만, 라몬만, 아산만, 울산만 등

- 피오르
 - 송네 피오르, 밀퍼드 사운드 등

6.8.3. 예문

- 반도
 - 황해연안에 돌출한 <변산반도: LCG_BAY>는 지괴산지의 대표적인 것으로
 - 따라서 대체로 <발칸반도: LCG_BAY>의 기후는 건조하며, 식생(植生)은 특히 남부에서 빈약하다.
 - 포르투갈은 1580년부터 에스파냐의 속령이 되었으며, 1640년에 독립을 회복한 후로는 이들 양국이 <이베리아반도: LCG_BAY>를 차지하고 있다.
- 곳
 - 올해 해돋이는 <호미곳: LCG_BAY>에서 보기로 했다.
- 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)는 노르웨이에서 가장 길고 수심이 깊은 피오르이다.

6.8.4. 주의 사항

- (태깅 대상 - 한반도) '한반도'는 국가와 지형을 모두 의미할 수 있으나 '한반도'에는 남한과 북한이 모두 있으므로 국가를 나타내는 LCP_COUNTRY로 태깅하지 않는다. 문맥상 지형명으로 사용된 경우만 LCG_BAY로 태깅한다. 지형명이 아닌 경우는 태깅하지 않는다.
 - 비행기가 한반도 상공을 지나고 있다. → 비행기가 <한반도: LCG_BAY> 상공을 지나고 있다.

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 기후구를 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_MOUNTAIN>, 북쪽에 <마등령: LCG_MOUNTAIN>·<미시령: LCG_MOUNTAIN> 등의 고개가 있다.

6.9.4. 주의 사항

- (태깅 대상 제외 - OO고개) '결딴 고개, 삼년 고개, 바람 고개' 등 정식 명칭이 아닌 명칭은 태깅하지 않는다.
- 넘어지면 삼 년 내에 죽는다고 알려진 삼년 고개에 관한 전설 → 넘어지면 삼 년 내에 죽는다고 알려진 삼년 고개에 관한 전설

6.10. LCG_ISLAND

6.10.1. 정의 및 설명

- 섬, 제도, 환초/산호섬의 명칭

6.10.2. 분석 대상

- 섬
 - 대난지도, 대부도, 백령도 등
 - 세빛섬, 서래섬, 나이팅게일섬, 노시베섬, 로드리게스섬 등
- 제도
 - 로스 제도, 마스카렌 제도 등

- 환초/산호섬
 - 윗제환초, 아리산호섬 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.
- 환초/산호섬
 - <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.

6.10.4. 주의 사항

- (태깅 범주 - LCG_ISLAND / LCP_PROVINCE) '제주도, 홋카이도, 오키나와, 코스라에주'와 같이 섬이면서 행정구역일 경우 LCP_PROVINCE로 우선 태깅한다. 문맥상 확실히 섬일 경우만 LCG_ISLAND로 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.
 - 코스라에주는 현재 전통 농업과 어업이 주를 이룬다. → <코스라에주: LCP_PROVINCE>는 현재 전통 농업과 어업이 주를 이룬다.

6.11. LCG_CONTINENT

6.11.1. 정의 및 설명

- 대륙 및 대륙의 일부분의 명칭

6.11.2. 분석 대상

- 대륙
 - 아메리카, 아시아, 오세아니아, 유라시아, 유럽, 북아메리카, 남아메리카, 아프리카 등
- 대륙의 일부분
 - 시베리아, 중동

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.
- 대륙의 일부분
 - <시베리아: LCG_CONTINENT>의 추위는 상상을 초월한다.
 - <중동: LCG_CONTINENT>지역 시장 점유율 15% 달성

6.11.4. 주의 사항

- (태깅 대상 - 줄임말) '동남아, 북미/남미' 등 대륙의 일부를 지칭하는 명칭이 줄임말로 사용되었을 경우도 LCG_CONTINENT로 태깅한다.
 - 동남아는 뭉치뭉치해도 리조트 콕 아니겠냐며 → <동남아: LCG_CONTINENT>는 뭉치뭉치해도 리조트 콕 아니겠냐며
- (태깅 범주 - 남극/북극) '남극/북극'의 경우 '대륙'을 나타내는 경우 LCG_CONTINENT로 태깅하고 '남극점/북극점'을 나타내면 LC_OTHERS로 태깅한다. 문맥으로 판단하기 어려운 경우 LCG_CONTINENT로 우선 태깅한다.
 - 일반인이 남극을 체험하기 위해서는 외교부 장관의 승인을 받아야 한다. → 일반인이 <남극: LCG_CONTINENT>을 체험하기 위해서는 외교부 장관의 승인을 받아야 한다.
- (태깅 범주 - LCG_CONTINENT / LC_OTHERS) '북아시아 지역'으로 정의되어 있는 '시베리아', '서아시아 일대'로 정의되어 있는 '중동'은 아시아(대륙)의 일부분이기 때문에 LCG_CONTINENT로 태깅한다. '아랍'은 특정 대륙의 일부분이 아닌 특정 지역에 해당하기 때문에 LCG_CONTINENT로 태깅하지 않고 LC_OTHERS로 태깅한다. '지방, 지역'은 같이 태깅하지 않는다.
 - <시베리아: LCG_CONTINENT>의 추위는 상상을 초월한다.
 - <중동: LCG_CONTINENT>지역 시장 점유율 15% 달성
 - '중동'만 LCG_CONTINENT로 태깅한다.
 - <아랍: LC_OTHERS>에 투자

6.12. LC_SPACE

6.12.1. 정의 및 설명

- 항성, 행성의 명칭
- 위성/유성 및 혜성의 명칭
- 별자리, 성운, 달의 명칭

6.12.2. 분석 대상

- 항성, 행성
 - 태양, 해, 달, 지구, 수성, 금성, 토성, 명왕성 등
- 위성/유성 및 혜성
 - 미란다, 거문고자리 유성군, 단주기혜성, 모어하우스혜성, 헬리혜성 등
- 별자리, 성운, 달
 - 사자자리, 게자리, 나침반자리, 물병자리, 오리온자리, 고리성운, 견우성, 게성운, 슈퍼문, 반달, 보름달 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
 - <미란다: LC_SPACE>를 비롯한 <천왕성: LC_SPACE>의 위성들도 <천왕성: LC_SPACE>의 자전면과 거의 일치하는 궤도를 가지고 있다.
 - 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, 달
 - 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.
 - 보통 바다 수위는 <슈퍼문: LC_SPACE>이 나타날 때 가장 높다.

6.12.4. 주의 사항

- (태깅 대상 제외 - 보통명사) '블랙홀, 별'과 같이 특정한 개체를 뜻하지 않고 보통명사로 쓰인 것은 태깅하지 않는다. '북극성'과 같이 특정 별의 이름을 뜻할 경우만 LC_SPACE로 태깅한다.
 - 별 중에서 북극성 → 별 중에서 <북극성: LC_SPACE>

6.13. LC_OTHERS

6.13.1. 정의 및 설명

- LC 계열의 세부 유형에 속하지 않는 기타 장소의 명칭
 - 지하철/기차역, 공항, 버스터미널, IC, 정류장/정거장/주차장, 시장, 마을 등의 구체적인 명칭
 - 농장, 야영장, 체험장, 수련장, 해수욕장, 스키장, 슬로프 등의 구체적인 명칭
 - 식물원, 수목원, 과수원/농원, 숲, 목장, 정원, 공원 등의 구체적인 명칭
 - 세트장, 광장, 테마파크/놀이공원 등의 구체적인 명칭
 - 항구/항만/포구, 골짜기, 해안/해변, 온천, 간헐천, 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, 검문소, 쇼핑/경제/산업/호텔/관광/아파트/공업 단지 등의 구체적인 명칭
 - 특정한 지역/권역/구역을 나타내는 명칭
 - 신도시, 기업도시 등 도시의 특정 구역을 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, 스키장, 슬로프 등의 구체적인 명칭
 - 노안청소년야영장, 새마을주말농장, 고현 어촌 체험 마을, 새내역사체험마을, 예송리해수욕장, 이스턴 슬로프, 루키 슬로프 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/놀이공원 등의 구체적인 명칭
 - 간현유원지, 대성산유원지, 5·18민주광장, 광화문 광장, 송상현 광장, 에버랜드, 디즈니월드, 디즈니랜드, 서울랜드 등
- 항구/항만/포구, 골짜기, 해안/해변, 온천, 간헐천, 폭포, 계곡, 동굴, 샘터, 약수터 등의 구체적인 명칭
 - 부산항, 인천항, 소래포구, 장림포구, 수통골 등
 - 해운대(해안), 용머리 해안, 닭머리 해안, 양남해변, 백암온천, 올드페이스풀 간헐천, 용정약수터 등
- 언덕/오름, 구릉, 나루, 금광, 사구, 발전소 등의 구체적인 명칭
- 가로구릉, 기아나고지, 나가구릉, 구드래 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, 검문소, 쇼핑/경제/산업/호텔/관광/아파트/공업 단지 등의 구체적인 명칭
 - 금강 휴게소, 죽전 휴게소, 장터목대피소, 소청대피소, 배둔검문소 등
 - 롯데몰, 영통/파주 프리미엄아울렛, 스타필드, 월스트리트, 천안산업단지, 성서산업단지, 보르시벨트, 주공5단지, 길음뉴타운 8단지, 중문관광단지, 보문관광단지, 구로공단 등
- 특정한 지역/권역/구역을 나타내는 명칭

- 서해안/동해안/남해안, 서양/동양, 중화권/이슬람권, 아랍, 동구/서구, 북방/서방/남방/동방, 적도, 계토 등
- 신도시, 기업도시 등 도시의 특정 구역을 나타내는 명칭
 - 서충주 신도시, 동탄 신도시, 판교 신도시, 판교지구, 판교테크노밸리, 원주기업도시 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 #<자갈치시장: LC_OTHERS> #물회
 - 축제의 주 행사인 산행은 오전 9시 명석면 <홍지마을 주차장: LC_OTHERS>에서 모여 준비 운동으로 몸을 풀고 출발하며, 해발 420m의 <광제봉수대: LC_OTHERS>에 올라 주변 경관을 감상한 뒤 <홍지마을: LC_OTHERS>까지 되돌아오는 13구간에서 실시된다.
- 농장, 야영장, 체험장, 수련장, 해수욕장, 스키장, 슬로프 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 현재 총 8개의 해수욕장이 개장되어 있다.
 - 12월 10일 화요일 <이스턴슬로프: LC_OTHERS>가 추가 오픈됩니다.
- 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, 간헐천, 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.
- 미국 <엘로스톤 국립공원: LC_OTHERS>의 엘로스톤 칼데라 안에 있는 <올드페이스폴 간헐천: LC_OTHERS>은 40분에서 80분 간격으로 분출한다.
- 언덕/오름, 구릉, 나루, 금광, 사구, 발전소 등의 구체적인 명칭
 - 부소산성 서쪽으로 백마강가의 <구드래나루: LC_OTHERS>는 지금은 공원으로 꾸며져 있다.
 - 이번 숙소는 <셋별오름: 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 300km²의 오아시스로 대추야자를 재배한다.
 - 고창군에 위치한 갯벌을 <고창 갯벌: LC_OTHERS>(내만형 갯벌)이라고 합니다.
- 휴게소, 대피소, 검문소, 쇼핑/경제/산업/호텔/아파트/관광/공업 단지 등의 구체적인 명칭
 - 봉정암에서 <소청대피소: LC_OTHERS>를 거쳐 대청봉까지
 - 강모씨가 승용차 운전 중 고성군 <배둔검문소: LC_OTHERS> 앞 도로에 떨어져 있던 철구조물과 부딪혀 승용차가 파손되었다.
 - 드디어 <죽전 휴게소: LC_OTHERS>, 곧 서울이다.
 - 간만에 쇼핑물 음 #<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 즐길거리 찾기
- 특정한 지역/권역/구역을 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.
- 6세기 경 콘스탄티노플에는 서부 유럽의 도시들보다 한참 전에 <게토: LC_OTHERS>가 있었다.
- <아랍: LC_OTHERS>에 투자
- 신도시, 기업도시 등 도시의 특정 구역을 나타내는 명칭
 - <동탄신도시: LC_OTHERS> 개발 논란
 - <판교테크노밸리: LC_OTHERS> 입주기업 행사

6.13.4. 주의 사항

- (태깅 대상 제외 - 보통명사) '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_RIVER>은 정기적으로 범람하는 강이다.
 - 여름 휴가는 산보다 바다가 낫지. 올해는 태안 반도? → 여름 휴가는 산보다 바다가 낫지. 올해는 <태안 반도: LCG_BAY>?
 - 이 근처에서 제일 가까운 놀이동산이나 동물원이 어디야? → 이 근처에서 제일 가까운 놀이동산이나 동물원이 어디야?
 - 특정한 것을 지칭하는 '이집트, 나일강, 태안 반도'는 LC로 태깅하지만 '강, 산, 바다, 놀이공원, 동물원'과 같이 특정한 것을 지칭하지 않는 보통명사는 태깅하지 않는다.
- (태깅 대상 제외 - OO바위) '울산바위, 갯바위'와 같은 OO바위는 특정한 공간의 의미가 아니므로 태깅하지 않는다.
 - 속초를 병풍처럼 감싸고 있는 설악산을 바라보면 전면에 보이는 바위가 바로 울산바위다. → 속초를 병풍처럼 감싸고 있는 설악산을 바라보면 전면에 보이는 바위가 바로 울산바위다.
- (태깅 대상 제외 - OO꽃밭) 'OO꽃밭'인 경우 구로 구성되어 있으므로 분리가 가능하기 때문에 'OO꽃밭' 전체를 LC_OTHERS로 태깅하지 않는다. 'OO꽃밭'의 OO만 내부 태깅하고 '꽃밭'은 태깅하지 않는다.
 - 양귀비 꽃밭 → <양귀비: PT_FLOWER> 꽃밭
- (태깅 대상 제외 - OO밭) 'OO밭'의 경우 특정 장소를 가리키는 것이 아니기 때문에 태깅하지 않는다. 또한 'OO밭'은 내부 태깅도 하지 않는다. 'OO밭'은 표기 유형은 같으나 사전 등재 여부가 다르다. 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
- (태깅 대상 제외 - OO방) '목욕탕, 노래방, PC방, 찜질방' 등을 특정 장소로 볼 수도 있지만 '식당, 카페 명'을 OGG_FOOD로 태깅한 것과 비슷하게 LC_OTHERS로 태깅하지 않는다. 'OO목욕탕, OOn노래방,

OOPC방, OO찜질방' 등 특정한 명칭이 있는 경우 OGG_ECONOMY로 태깅한다.

- 시험이 끝난 후 PC방에서 모였다. → 시험이 끝난 후 PC방에서 모였다.
 - PC방이 특정한 장소를 뜻하지 않으므로 태깅하지 않는다.
- 시험이 끝난 후 수노래방에서 모였다. → 시험이 끝난 후 <수노래방: OGG_ECONOMY>에서 모였다.
 - '수노래방'과 같이 특정한 명칭이 있는 경우 OGG_ECONOMY로 태깅한다.
- (태깅 대상 제외 - OORoom, OORoom, OORoom) 'OORoom, OORoom, OORoom'을 건물 내 특정 장소로 볼 수도 있지만 건물에 포함되어 있는 것으로 보고 태깅하지 않는다. '현대백화점 VIP룸', '세종문화회관 예인홀', '경기대학교 체육관'과 같은 경우 '현대백화점'이나 '세종문화회관', '경기대학교'만 태깅하고, 백화점/회관/건물/대학교 안의 장소인 'OORoom, OORoom, OORoom' 등은 태깅하지 않는다.
- 경기대학교 체육관 → <경기대학교: OGG_EDUCATION> 체육관
- (태깅 대상 제외 - OO전망대) '통일전망대, 앞산공원전망대'와 같은 'OO전망대'의 경우 인공물, 건축물이기 때문에 LC_OTHERS가 아니라 AF_BUILDING으로 태깅한다.
 - <통일전망대: AF_BUILDING>는 다소 아득하게 느껴졌던 분단의 현실을 직시할 수 있는 곳이다.
- (태깅 단위 - OO신도시, OO지구 등) '신도시, 지구, 테크노밸리' 등은 앞에 지역명이 같이 나올 때만 묶어서 LC_OTHERS로 태깅한다. '신도시, 지구, 테크노밸리' 등이 단독으로 나왔을 경우 태깅하지 않는다.
 - 이번에 동탄 신도시로 이사를 갔다. → 이번에 <동탄 신도시: LC_OTHERS>로 이사를 갔다.
 - '신도시' 앞에 지역명인 '동탄'이 있으므로 '동탄 신도시'를 묶어서 LC_OTHERS로 태깅한다.
 - 신도시 땅을 사 놓아야 한다. → 신도시 땅을 사 놓아야 한다.
 - '신도시' 앞에 지역명이 없이 보통명사로 쓰였기 때문에 태깅하지 않는다.
 - 친구를 만나러 동탄에 갔다. → 친구를 만나러 <동탄: LCP_COUNTY>에 갔다.
 - '신도시'가 없이 지역명만 나왔을 경우, 해당 지역명이 속하는 행정구역명으로 태깅 범주를 판단한다. '동탄'은 '동'에 해당하므로 LCP_COUNTY로 태깅한다.
- (태깅 범주 - LC_OTHERS / AF_CULTURAL_ASSET) 문화재로 등재되어 있지 않으면 LC_OTHERS로 태깅하고, '전주 객사'와 같이 구체적인 유적지, 마을, 터, 성지, 전승지 등이 문화재로 등록되어 있으면 AF_CULTURAL_ASSET로 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.
 - '토성민속마을'은 문화재로 등록되어 있지 않기 때문에 LC_OTHERS로 태깅하고 '아산외암민속마을'은 문화재로 등록되어 있으므로 AF_CULTURAL_ASSET으로 태깅한다.
- (태깅 범주 - 경포대) '경포대'는 문맥에 따라 해변, 건물 등으로 모두 태깅 가능하다. 이때 문맥과 적정 술어로 판정할 수 있다. '봄비다, 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.
 - 문맥상 '해변'을 지칭하므로 LC_OTHERS로 태깅한다.
 - 살짝 보이는 누각, 강릉 경포대를 한눈에 담아 봅니다. 처마 부분이 정말 화려하지요? → 살짝 보이

는 누각, 강릉 <경포대: AF_BUILDING>를 한눈에 담아 봅니다. 처마 부분이 정말 화려하지요?

■ 문맥상 '누각'을 지칭하므로 AF_BUILDING으로 태깅한다.

- (태깅 범주 - 해운대) '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
 - 문맥상 해수욕장/해변을 뜻하므로 LC_OTHERS로 태깅한다.
 - 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.
 - 문맥상 '해운대'가 부산의 해운대구를 뜻하므로 LCP_COUNTY로 태깅한다.
- (태깅 범주 - LC_OTHERS / AF계열 / OGG계열) '서면롯데백화점, 현대백화점 판교'와 같이 특정한 'OO백화점'이나 '홍대/이대' 등 특정한 'OO대(학교)'와 같은 OGG계열이 '~에서 만나다/구입하다/팔다/사다' 등과 함께 사용되어 문맥상 '기관'을 의미하지 않고 '장소/위치'를 의미하는 경우 LC_OTHERS로 태깅한다. 단, OGG 계열이 나왔을 때 개별 상점의 경우 '~에 가다, ~에서 만나다' 등의 술어가 나왔다고 하더라도 OGG 계열로 태깅한다. 일정한 구역에 여러 상점이 한곳에 모여 있는 복합문화공간, 쇼핑물의 경우에만 한정하여 LC_OTHERS로 태깅한다. '~에 다니다, ~을 경영하다, ~이 흥하다, ~이 잘 되다, ~을 차리다'와 같은 술어와 함께 쓰이면 OGG로 태깅하고 '~이 보인다/흔들리다, ~을 짓다/만들다'와 함께 쓰이며 건축물의 의미로 해석되는 경우 AF로 태깅한다.
 - 어제 홍대에서 산 핑크색 거울 니트 → 어제 <홍대: LC_OTHERS>에서 산 핑크색 거울 니트
 - 서강대 근처에 있는 카페에서 → <서강대: LC_OTHERS> 근처에 있는 카페에서
 - 위 예문들의 '홍대, 서강대'는 학교 일대 지역을 뜻하므로 LC_OTHERS로 태깅한다.
 - 그는 광양제철고-중앙대를 거쳐 안산 그리너스에서 프로에 데뷔했다. → 그는 <광양제철고: OGG_EDUCATION>-<중앙대: OGG_EDUCATION>를 거쳐 안산 그리너스에서 프로에 데뷔했다.
 - 장소/위치가 아닌 졸업한 기관을 뜻하므로 OGG_EDUCATION으로 태깅한다.
 - 명동에 있는 <롯데백화점: AF_BUILDING>은 1979년에 지어졌다.
 - '기관'이 아닌 '건축물'의 개념으로 사용되었으므로 AF_BUILDING으로 태깅한다.
 - 퇴근 후에 다이소에 갔다. → <다이소: OGG_ECONOMY>에 갔다.
 - 스타벅스에 가서 산 텀블러. → <스타벅스 : OGG_FOOD>에 가서 산 텀블러.
 - CGV에 가서 봤던 영화가 → <CGV : OGG_ART>에 가서 봤던 영화가
 - '다이소, 스타벅스, CGV'가 '-에 가다'의 서술어와 쓰여 LC로 볼 수 있으나 OGG 계열의 개별 상점이므로 OGG로 태깅한다.
- (태깅 범주 - LC_OTHERS / 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.
- (태깅 범주 - LC_OTHERS / OGG_HOTEL) 휴양시설은 OGG_HOTEL과 LC_OTHERS가 모두 가능한데 '호텔, 모텔, 민박, 펜션, 콘도, 리조트, 산장'의 특정 명칭만 OGG_HOTEL로 태깅하고 그 외 휴양시설인 '야

영장, 체험장' 등의 명칭은 LC_OTHERS로 태깅한다.

- 이번 휴가는 <신라호텔: OGG_HOTEL>로 가기로 했다.
- <용연 청소년 야영장: LC_OTHERS>은 청소년들에게 건전한 야영 활동과 심신 수련 및 호연지기를 키워 주기 위하여 건립했다.
- (태깅 범주 - LC_OTHERS / LCP_COUNTY) '신촌, 강남'이 '신촌역, 강남역'과 같이 문맥상 '역'이 확실할 경우만 LC_OTHERS로 태깅한다. 지하철역이 아닐 경우, '신촌, 강남'은 도시 안의 일부 지역이면서 행정 구역명으로도 쓰이는 지역이기 때문에 LCP_COUNTY로 태깅한다. 문맥으로 판단이 어려운 경우 LCP_COUNTY로 우선 태깅한다. 한편, '서면, 광안리, 압구정로데오'의 경우 지하철역 이름으로 그 일대를 일컫기는 하지만 정확한 행정구역 명칭이 아니므로 LCP_COUNTY로 태깅하지 않고 LC_OTHERS로 태깅한다.
 - 신촌 다음 역에서 내리면 된다. → <신촌: LC_OTHERS> 다음 역에서 내리면 된다.
 - 맥락상 '신촌역'을 나타내므로 LC_OTHERS로 태깅한다.
 - 지난 주말에는 대학 동기들과 시간을 맞춰 <서면: LC_OTHERS> 맛집에 다녀왔는데요.
 - 지난 주말엔 친구들과 신촌에서 모였어요! → 지난 주말엔 친구들과 <신촌: LCP_COUNTY>에서 모였어요!
 - 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.
 - '신촌, 시부야, 혜화'는 도시 안의 일부 지역을 나타내므로 LCP_COUNTY로 태깅한다.
- (태깅 범주 - LC_OTHERS / LCG_CONTINENT) '아랍'은 특정 대륙의 일부분이 아닌 특정 지역에 해당하기 때문에 LCG_CONTINENT로 태깅하지 않고 LC_OTHERS로 태깅한다. 한편, '북아시아 지역'으로 정의되어 있는 '시베리아', '서아시아 일대'로 정의되어 있는 '중동'은 아시아(대륙)의 일부분이기 때문에 LCG_CONTINENT로 태깅한다. '지방, 지역'은 같이 태깅하지 않는다.
 - <아랍: LC_OTHERS>에 투자
 - <시베리아: LCG_CONTINENT>의 추위는 상상을 초월한다.
 - <중동: LCG_CONTINENT>지역 시장 점유율 15% 달성
 - '중동'만 LCG_CONTINENT로 태깅한다.
- (태깅 범주 - 남극/북극) '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.
 - 일반인이 남극을 체험하기 위해서는 외교부 장관의 승인을 받아야 한다. → 일반인이 <남극: LCG_CONTINENT>을 체험하기 위해서는 외교부 장관의 승인을 받아야 한다.

7. CIVILIZATION

- CIVILIZATION은 문명/문화와 관련된 용어들을 대상으로 한다.

분류 명칭	정의 및 설명
7.1. CV_CULTURE	- 문명, 문화 - 혁명/사회 변혁
7.2. CV_TRIBE	- 고인류 학명 - 민족/종족/인종 명칭 - 국가를 구성하는 국민의 지칭
7.3. CV_LANGUAGE	- 언어 명칭
7.4. CV_POLICY	- 제도 및 정책의 명칭
7.5. CV_LAW	- 법/법률 명칭 - 감사 활동, 특정 처분이나 원칙을 나타내는 법률 용어 - 법적 권리, 범죄 명칭, 형벌 명칭, 명령/칙령, 협정, 협약의 명칭
7.6. CV_CURRENCY	- 통화 명칭
7.7. CV_TAX	- 조세 명칭
7.8. CV_FUNDS	- 연금, 기금, 자금, 펀드 명칭
7.9. CV_ART	- 예술의 분류, 종류 : 문예, 음악, 미술, 춤, 영화, 드라마의 종류
7.10. CV_SPORTS	- 스포츠/레포츠/레저 명칭 및 종목 - 정식 스포츠에 해당하는 게임 종목 - 전통놀이/민속놀이의 종류
7.11. CV_SPORTS_POSITION	- 스포츠 포지션
7.12. CV_SPORTS_INST	- 스포츠/레포츠/레저의 용품/도구 명칭
7.13. CV_PRIZE	- 상, 훈장 명칭
7.14. CV_RELATION	- 가족/친족의 구성원 - 가족/친족 관계를 지칭하는 명사
7.15. CV_OCCUPATION	- 직업 명칭 - 적정 술어 : '~으로 일하다, ~을 해서 먹고 살다, ~을 해서 생계를 유지하다' 등
7.16. CV_POSITION	- 공무원/군인/교수의 직위/직책/직급, - 직장/학교/고대 국가에서의 직위/직책/직급
7.17. CV_FOOD	- 요리의 유형 : 국, 밥, 국수, 안주 - 요리명, 식품명 - 식재료
7.18. CV_DRINK	- 음료/주류의 하위 부류 - 음료/주류명

7.19. CV_FOOD_STYLE	- 음식 유형
7.20. CV_CLOTHING	- 섬유/의복의 명칭 - 의복의 유형, 모자, 가방, 양말, 신발의 하위 유형
7.21. CV_BUILDING_TYPE	- 건축 양식 명칭

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 <아즈텍문명: 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 막아주는 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 <도식문화: 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.

7.1.4. 주의 사항

- (태깅 대상 - 산업혁명, 르네상스) 시기, 기간을 특정할 수 없는 사조나 사회변화를 의미하는 OO혁명은 EV_WAR_REVOLUTION으로 태깅하지 않고 CV_CULTURE로 태깅한다.
- n차산업혁명, 르네상스
 - '르네상스'는 예술운동 외의 학문, 사상 등까지 다루고 있으므로 예술사조로 보기 힘들다. 즉, TR_ART로 태깅하지 않는다.

7.2. CV_TRIBE

7.2.1. 정의 및 설명

- 고인류 학명, 민족, 종족, 인종, 국가를 구성하는 국민의 지칭

7.2.2. 분석 대상

- 고인류 학명
 - 호모사피엔스, 호모사피엔스사피엔스, 호모에렉투스, 네안데르탈인 등
- 민족, 종족, 인종
 - 이라크족, 유대인, 조선족, 마오리족, 아메리카인디언, 아리아인, 고산족, 집시, 보헤미안, 유목민, 아프리카인, 아메리카인, 아시아인, 흑인, 백인, 황인 등
- 국가를 구성하는 국민의 지칭
 - 한국인, 미국인, 중국인, 일본인, 한국계, 유대인계, 히스패닉계 등

7.2.3. 예문

- 고인류 학명
 - <호모사피엔스: CV_TRIBE>는 슬기로운 사람이라는 뜻을 가지고 있다.
- 민족, 종족, 인종
 - <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 문화 수준도 종족 간에 격차가 심하다.
- 뉴욕 주방에는 <아시아인: CV_TRIBE>이 거의 없었다.
- 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 <아메리카인도인: CV_TRIBE>이라 부르게 되었다.
 - 이번 행사로 <한국계: CV_TRIBE> 상인들에 대한 선호도가 높아졌다.

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, 한국 수어, 프랑스 수어 등

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.
 - 프랑스 수어 → <프랑스 수어: CV_LANGUAGE>
 - 한국 수화 언어 → <한국 수화 언어: CV_LANGUAGE>

7.3.4. 주의 사항

- (태깅 대상 제외 - 보통명사) '외국어, 모국어' 등의 보통명사는 CV_LANGUAGE로 태깅하지 않는다. 개체명으로 보기 어렵다.
 - 외국어, 모국어, 표준어, 방언, 수화, 수어 등
 - 일본 수어 → <일본 수어: CV_LANGUAGE>
 - '일본 수어'는 특정한 언어를 지칭하므로 CV_LANGUAGE로 태깅한다.
- (태깅 대상 제외 - 방언) 하위 부류로 볼 수 있는 '동남 방언' 등은 모두 '한국어, 영어' 등과 같은 계열로 볼 수 없으므로 태깅 대상이 아니다.
- (태깅 범주 - CV_LANGUAGE / LCP_COUNTRY) '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.
 - 단순히 한국어로 된 드라마를 지칭하는 것이 아닌 한국에서 제작된 드라마를 지칭하므로 CV_LANGUAGE가 아닌 LCP_COUNTRY로 태깅한다.
 - <영어: CV_LANGUAGE> 버전으로 먼저 출간된 이 한국어 교재는 현재 <중국: CV_LANGUAGE> 버전과 <일본: CV_LANGUAGE> 버전의 번역 작업이 진행 중이다.
 - '일본'이 확실히 특정 언어들을 지칭하므로 CV_LANGUGAE로 태깅한다.
- (태깅 범주 - CV_LANGUAGE / FD_HUMANITIES) '영어, 국어'와 같은 언어의 경우, '-로 말하다, -로 이야기하다'와 같은 술어와 같이 쓰일 경우에는 CV_LANGUAGE로 태깅하고, '-을 공부하다, -을 연구하다, -을 분석하다'와 같은 술어와 함께 쓰이거나 '수능, 과목, 난이도'와 같은 단어와 어울려 쓰일 경우에는 FD_HUMANITIES로 태깅한다.
 - 제니는 영어로 유창하게 말한다. → 제니는 <영어: CV_LANGUAGE>로 유창하게 말한다.
 - 위 문장에서 '영어'는 언어를 뜻하는 것이기 때문에 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.

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, 아동수당, 주택보조금, 무상급식, 계엄령, 비상계

업, 경비계업 등

7.4.3. 예문

- 제도 및 정책의 명칭
 - <가부장제: CV_POLICY>의 가족형태에서는 가족성원이 세습적 규칙에 따라 지명된 개인의 지배를 받는데, 대개는 장남이 세습적으로 가정의 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.
 - 내년부터 <무상급식: CV_POLICY>이 전격 시행된다.

7.4.4. 주의 사항

- (태깅 대상 제외 - 광범위한 의미의 정책/제도) '교육복지 정책, 결혼 제도'는 세부적인 제도, 정책으로 볼 수 없기 때문에 태깅하지 않는다.
 - 교육복지 정책, 결혼 제도 → 교육복지 정책, 결혼 제도
 - 세부적인 제도를 칭하지 않기 때문에 태깅하지 않는다.
 - 무상급식 제도 시행 → <무상급식 제도: CV_POLICY> 시행
 - '무상급식 제도'는 특정한 제도이기 때문에 CV_POLICY로 태깅한다.
- (태깅 대상 제외 - OO회의) '제가회의, 화백회의, 정사암회의 등'은 과거 삼국시대의 귀족 회의로 제도 및 정책이 아니므로 CV_POLICY로 태깅하지 않는다.
- (태깅 대상 제외 - 국가 사업) 국가 사업은 정책, 제도에 한해서 CV_POLICY로 태깅한다. 그 외 국가 사업은 TMI_PROJECT로 태깅한다.
 - 화폐정리사업 → <화폐정리사업: TMI_PROJECT>
 - '화폐정리사업'은 국가 사업이므로 TMI_PROJECT로 태깅한다.

7.5. CV_LAW

7.5.1. 정의 및 설명

- 법/법률 명칭, 감사 활동, 특정 처분이나 원칙을 나타내는 법률 용어, 법적 권리, 범죄 명칭, 형벌 명칭, 명령/칙령, 협정, 협약의 명칭

7.5.2. 분석 대상

- 법/법률 명칭
 - 국제연합헌장, 민법, 형법, 함무라비법전, 관습법 등
 - 헌법, 국회법, 법원조직법 등
 - 민식이법, 김영란법 등
 - 대동법
- 감사 활동, 특정 처분이나 원칙을 나타내는 법률 용어
 - 국정감사, 불기소처분, 일사부재의 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
- 명령/칙령
 - 방곡령, 대사령, 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
 - 조선후기에 대동법을 시행하였다. → 조선후기에 <대동법: CV_LAW>을 시행하였다.
- 감사 활동, 특정 처분이나 원칙을 나타내는 법률 용어
 - 검사가 사건을 불기소처분을 하는 경우 → 검사가 사건을 <불기소처분: CV_LAW>을 하는 경우
- 법적 권리
 - 대의원 자격 요건은, 국회의원 <피선거권: CV_LAW>이 있는 30세 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.
- 명령/칙령
 - 국가에 경사가 있을 때 대사령을 내려 죄인들을 사면해 주었다. → 국가에 경사가 있을 때 <대사령: CV_LAW>을 내려 죄인들을 사면해 주었다.
- 협정, 협약의 명칭
 - 런던 협약은 비행기나 선박에서 나오는 쓰레기 투기를 규제하기 위한 협약이다. → <런던 협약: CV_LAW>은 비행기나 선박에서 나오는 쓰레기 투기를 규제하기 위한 협약이다.

7.5.4. 주의 사항

- (태깅 대상 - 형벌) 학교나 회사, 종교 내에서 만들어진 규율이나 형벌은 태깅하지 않는다.
- (태깅 대상 - 벌금) '벌금'은 법적 효력이 있는(벌금형 선고) 경우에만 태깅하며, 한 사회나 공동체에서 규칙으로 정한 벌금의 경우에는 태깅하지 않는다.
- 판사는 피고에게 벌금을 내라는 판결을 내렸다. → 판사는 피고에게 <벌금: CV_LAW>을 내라는 판결

을 내렸다.

- 판사가 선고한 것이므로 CV_LAW로 태깅한다.
- 철수는 지각으로 낸 벌금만 10만 원이다. → 철수는 지각으로 낸 벌금만 10만 원이다.
 - 이 문장에서 '벌금'은 법적 효력이 있는 선고가 아니라 소속 공동체의 규칙에 의해 낸 것이므로 태깅하지 않는다.
- (태깅 대상 제외 - 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
- (태깅 대상 제외 - 계엄) 계엄령, 비상계엄, 경비계엄 등과 같이 실제 시행이 되는 정책 용어인 경우는 법률 용어이지만 CV_LAW로 태깅하지 않고 CV_POLICY로 태깅한다.
 - 계엄령, 비상계엄, 경비계엄 → 모두 CV_POLICY로 태깅
- (태깅 단위 - 벌금, 처벌) '벌금 200만 원, 징역 12개월'과 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>
- (태깅 단위 - 법률 조항) 법률 조항의 경우도 각각 분리하여 태깅한다.
 - OO법 제3조 제 4항 → <OO법: CV_LAW>, <제3조: QT_ORDER> <제 4항: QT_ORDER>
- (태깅 범주 - OO조약, OO협정) '~조약', '~협정'은 문맥에 준하여 태깅한다. '-에 쓰여져 있다, -에 서명하다' 등 조약서, 협정서를 의미하는 경우는 CV_LAW로 태깅한다. '-을 체결하다, -을 맺다' 등의 적정 술어와 공기하는 경우이면 EV_OTHERS로 태깅한다. 문맥에서 판단하기 어려울 경우에는 EV_OTHERS를 우선순위로 정한다.
 - <포츠머스 조약: CV_LAW>에는 러시아가 조선으로부터 철수하고 모든 영향력을 내려 놓겠다는 내용이 쓰여 있다.
 - '쓰여 있다'와 공기하므로 CV_LAW로 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.
 - '체결하다, 맺다'와 공기하므로 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,

7.6.4. 주의 사항

- (태깅 대상 제외 - 금화, 은화) '금화, 은화'와 같이 특정 통화 명칭이 아닌 것은 보통명사로 보고 태깅하지 않는다.
 - 현대의 금화의 예로는 노르웨이의 크로네가 있다. → 현대의 금화의 예로는 노르웨이의 <크로네: CV_CURRENCY>가 있다.
 - '금화'는 보통명사기 때문에 태깅하지 않으며, 구체적인 통화 명칭인 '크로네'만 CV_CURRENCY로 태깅한다.
- (태깅 단위 - OO화) 우리말샘에 '달러화'가 복합어로 되어 있으므로 '달러화, 유로화'와 같이 '특정 통화 명칭+화'의 경우 전체를 CV_CURRENCY로 태깅한다.
 - <달러화: CV_CURRENCY>, <유로화: CV_CURRENCY>

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, 노령연금, 개인연금, 퇴직연금, 문예진흥기금, 통화안정기금, 방송발전기금, 공적자금 등

7.8.3. 예문

- 연금/기금/자금/펀드 명칭
 - 가입 기간이 20년 이상인 가입자 또는 가입자였던 자가 60세(대통령령이 정하는 특수직종근로자는 55세)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

- (태깅 대상 제외 - 금융 상품명) 금융 상품명인 경우, AFW_SERVICE_PRODUCTS로 태깅한다.
 - 높은 수익률을 원하신다면 <필승코리아증권투자신탁: AFW_SERVICE_PRODUCTS>을 추천드립니다.
 - 퇴직연금을 하실 거면 '퇴직연금 고배당주 증권투자신탁1호'를 추천드립니다. → <퇴직연금: CV_FUNDS>을 하실 거면 <퇴직연금 고배당주 증권투자신탁1호: AFW_SERVICE_PRODUCTS>를 추천드립니다.

- '퇴직연금'은 연금 종류의 명칭이므로 CV_FUNDS로 태깅하고 '퇴직연금 고배당주 증권자투자신탁 1호'는 구체적 금융 상품명이므로 AFW_SERVICE_PRODUCTS로 태깅한다.

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, 종류
 - 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, 케이팝
 - 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.
 - 푹푹한 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

- (태깅 대상 - 판소리, 종묘제례악) 유네스코 지정 한국의 무형유산인 '판소리'나 '종묘제례악'은 음악의 하위 장르로 볼 수 있으므로 CV_ART로 태깅하고, 구체적인 작품명은 AFA_MUSIC으로 태깅한다.
 - 국립창극단에서는 매년 판소리 수궁가가 공연된다. → 국립창극단에서는 매년 <판소리: CV_ART> <수궁가: AFA_MUSIC>가 공연된다.
- (태깅 대상 제외 - 상위 부류명) '문예, 음악, 미술, 춤, 영화, 드라마'와 같은 상위 부류는 태깅하지 않는다. '소설'은 문예 종류의 하나이면서 상위 부류명이므로 예외적으로 태깅한다.
 - 문예 → 문예

- 소설, 고전소설 → <소설: CV_ART>, <고전소설: CV_ART>
- (태깅 대상 제외 - 보통명사) '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 초대가 됐습니다. → 인기 절정의 <가요: CV_ART>보다는 음악인의 냄새가 담긴 그런 음악을 사랑하는 음악창고에 제가 오늘 초대가 됐습니다.
 - '음악'은 보통명사로 태깅하지 않는다. 세부 유형인 '가요'만 CV_ART로 태깅한다.
- (태깅 대상 제외 - 예술 관련 학문) '문예학, 음악사학'과 같은 예술 관련 학문 분야는 FD_ART로 태깅하고, '고전소설, 소나타'와 같은 예술의 한 분야인 경우 CV_ART로 태깅한다.
 - 문예학을 전공하면서 고전소설을 많이 접했다. → <문예학: FD_ART>을 전공하면서 <고전소설: CV_ART>을 많이 접했다.
 - '문예학'은 적정 술어 '전공하다'와 같이 쓰였기 때문에 학문으로 보고 FD_ART로 태깅하고 예술의 한 분야인 '고전소설'은 CV_ART로 태깅한다.
- (태깅 대상 제외 - 특정 작품명) 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.
 - '라틴아메리카댄스'는 춤의 종류이기 때문에 CV_ART로 태깅한다.
- (태깅 대상 제외 - 결과물) 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 낚두리했다.
 - '만용에게'는 작품명이므로 AFA_DOCUMENT로 태깅한다.
- (태깅 단위 - 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>

7.10. CV_SPORTS

7.10.1. 정의 및 설명

- 스포츠/레포트/레저 명칭 및 종목, 정식 스포츠에 해당하는 게임 종목, 전통놀이/민속놀이의 종류

7.10.2. 분석 대상

- 스포츠/레포트/레저 명칭 및 종목

- 스케이트, 스키, 하이킹, 행글라이딩, 인라인스케이팅, 축구, 마라톤 등
- 정식 스포츠에 해당하는 게임 종목
 - 바둑, 당구, 포켓볼 등
- 전통놀이/민속놀이의 종류
 - 널뛰기, 연날리기, 칠교놀이, 돛다리밟기, 사물놀이, 윷놀이, 자치기, 제기차기 등

7.10.3. 예문

- 스포츠/레포츠/레저 명칭 및 종목
 - 일상생활에서 교외로 벗어나는 산책 나들이에서부터 가벼운 등산 등 야외활동을 널리 <하이킹: CV_SPORTS>이라 한다.
 - <행글라이딩: CV_SPORTS>은 공기 속을 움직이는 스포츠이므로 기상에 민감한 편인데 기상조건으로 는 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.
- 정식 스포츠에 해당하는 게임 종목
 - <당구: CV_SPORTS> 경기의 일종인 <포켓볼: CV_SPORTS>
- 전통놀이/민속놀이의 종류
 - 대보름이면 가족들이 함께 <연날리기: CV_SPORTS>를 했다.

7.10.4. 주의 사항

- (태깅 대상 제외 - 일반적인 놀이, 게임) 전통놀이, 민속놀이의 종류가 아닌 달리기, 줄넘기, 공놀이, 공기놀이 등은 태깅하지 않는다. 마찬가지로 게임으로만 행해지며 전문 종목, 스포츠로 볼 수 없는 경우는 태깅하지 않는다.(태깅 대상 제외 항목을 묶음)
 - 설날의 놀이에는 윷놀이, 널뛰기가 있다. → 설날의 놀이에는 <윷놀이: CV_SPORTS>, <널뛰기: CV_SPORTS>가 있다.
 - '놀이'는 일반 명사로 태깅하지 않고, 전통놀이, 민속놀이의 종류인 '윷놀이, 널뛰기'만 CV_SPORTS로 태깅한다.
 - 어휘 학습을 위해 빙고게임을 했다. → 어휘 학습을 위해 빙고게임을 했다.
 - '빙고게임'은 전문 종목이나 정식 스포츠로 행해지는 게임이 아니기 때문에 CV_SPORTS로 태깅하지 않는다.

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

- 스포츠 포지션
 - <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.
 - 동양에서는 전통적으로 최고의 <타자: CV_SPORTS_POSITION>를 <4번 타자: CV_SPORTS_POSITION>에 놓는다.
 - 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.

7.12. CV_SPORTS_INST

7.12.1. 정의 및 설명

- 스포츠/레포트/레저의 용품/도구 명칭

7.12.2. 분석 대상

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

7.12.3. 예문

- 스포츠/레포트/레저의 용품/도구 명칭

- 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

- (태깅 대상 제외 - 운동 의복) 수영복, 등산복 등의 운동 의복은 CV_SPORTS_INST로 태깅하지 않는다. CV_CLOTHING으로 태깅한다.
- 테니스 교실에 테니스채와 테니스복을 챙겨 갔다. → 테니스 교실에 <테니스 채: CV_SPORTS_INST>와 <테니스복: CV_CLOTHING>을 챙겨 갔다.
 - 테니스 도구인 '테니스 채'는 CV_SPORTS_INST로 태깅하고, '테니스복'은 옷의 종류로 CV_CLOTHING으로 태깅한다.
- (태깅 대상 제외 - 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.
- <버터플라이: OGG_ECONOMY>에서 <탁구채: CV_SPORTS_INST>는 <티모볼: AFW_OTHER_PRODUCTS>이나 <니추고: AFW_OTHER_PRODUCTS>가 가성비가 좋다.
 - '탁구채'는 스포츠 용품이므로 CV_SPORTS_INST로 태깅하고, 탁구채상품명인 '티모볼, 니추고'는 AFW_OTHER_PRODUCTS로 태깅한다. '버터플라이'는 회사명이므로 OGG_ECONOMY로 태깅한다.
- (태깅 대상 제외 - 운송수단) '자전거, 요트'와 같은 운송수단은 CV_SPORTS_INST로 태깅하지 않고 AF_TRANSPORT로 태깅한다.

7.13. CV_PRIZE

7.13.1. 정의 및 설명

- 상과 훈장

7.13.2. 분석 대상

- 상과 훈장
 - 대상, 최우수상, 최고상, 우수상, 인기상
 - 노벨경제학상, 노벨문학상, 노벨물리학상, 노벨생리의학상, 노벨평화상, 노벨화학상 등
 - 대통령상, 경제정의기업상, 자기상표제품 수출상, 대중상

7.13.3. 예문

- 상과 훈장
 - <대중상: CV_PRIZE>은 한국영화계의 대표적 영화제로 존속하면서 우수영화의 발굴 및 포상을 통한 진흥효과를 거둔 것은 긍정적이지만

7.13.4. 주의 사항

- (태깅 대상 제외 - 우승) '대상, 최우수상, 최고상, 우수상' 등의 보통명사는 CV_PRIZE로 태깅한다. 그러나 '우승, 준우승'은 CV_PRIZE로 태깅하지 않는다.
 - 대상 → <대상: CV_PRIZE>
 - 우승 → 우승
- (태깅 단위 - 날짜+순서+상) 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

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, 사촌 동생, 사촌 오빠, 예비신랑, 예비신부
- 가족/친족 관계를 지칭하는 명사
 - 부부, 부모, 형제, 자매, 남매, 부자, 모녀, 친척, 사촌 형제, 학부모, 배필
 - 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
 - 내 <예비신랑: CV_RELATION>을 소개해 볼까 해요.
- 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

- (태깅 대상 - 가족, 아이) '가족, 아이'의 경우, 가족의 구성원이나 가족 관계를 나타낼 때에는 CV_RELATION으로 태깅한다. '음악의 아버지'와 같이 가족/친족의 대상이 아니거나 보통명사로 쓰인 경우는 CV_RELATION으로 태깅하지 않는다.
 - 너의 가족에게도 통지가 갔을 것이었다. → 너의 <가족: CV_RELATION>에게도 통지가 갔을 것이었다.
 - 이 문장에서 '가족'은 가족의 구성원을 칭하므로 CV_RELATION으로 태깅한다.
 - 전시회에 가족 단위의 관람객들이 줄을 잇고 있다.
 - '가족'이 가족 관계를 나타내지 않고 보통명사로 쓰였으므로 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.
 - '아이'가 '우리집'의 가족 구성원을 칭하므로 CV_RELATION으로 태깅한다.
 - 먹는 것을 가지고 어른이 아이를 울리면 어떻게 해요?
 - '어른, 아이'가 가족 관계를 나타내지 않으므로 CV_RELATION으로 태깅하지 않는다.
 - 바흐가 음악의 아버지라고 불리는 이유는 바로크 시대 이전의 음악이 오늘날의 서양 음악으로 발전할 수 있도록 음악의 기초를 닦았기 때문이지.
 - 이 문장에서 '아버지'는 가족 관계를 나타내는 게 아니라 비유를 위해 쓰였으므로 CV_RELATION으로 태깅하지 않는다.
- (태깅 단위 - 예비신부, 예비신랑) 예비신랑, 예비신부의 경우 가족 구성원의 의미를 나타낸다고 볼 수 있으므로 전체를 CV_RELATION으로 태깅한다.
 - 내 예비신랑을 소개해 볼까 해요. → 내 <예비신랑: CV_RELATION>을 소개해 볼까 해요.
 - 예비신부님, 들어오십니다. → <예비신부: CV_RELATION>님, 들어오십니다.
- (태깅 단위 - 사촌 OO) '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>

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, 교황, 수녀, 추기경, 수도사, 전도사, 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
 - 육군, 공군, 해군, 의병 등
- 기타
 - 학생, 탐정, 큐레이터, 나레이터모델, 외환딜러, 브로커, 스파이, 오피상, 마도로스, 장돌뱅이, 기공, 기수(騎手), 유튜버, 블로거

7.15.3. 예문

- 직업
 - 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

- (태깅 대상 - 학생) '학생'은 위의 적정 술어와 어울리지 않지만 '직업란'에 '학생'을 쓰는 경우가 있으므로 예외적으로 CV_OCCUPATION으로 태깅한다. 이외에 학생의 하위 분류인 '대학생, 고등학생, 중학생, 초등학생' 등은 태깅 대상이 아니다.
 - 제 직업은 학생입니다. → 제 직업은 <학생: CV_OCCUPATION>입니다.
 - 대학생이세요? → 대학생이세요?
 - '학생'은 CV_OCCUPATION으로 태깅하지만 '대학생'은 학생의 하위 분류 대상이므로 태깅하지 않는다.
- (태깅 대상 - 선생님) '교사'와 같이 직업을 나타낼 경우만, CV_OCCUPATION으로 태깅한다. 직업이 아닌 남을 존대하여 쓰인 '선생님'은 태깅 대상이 아니다.
 - 김 씨는 선생님이로 일하고 있다. → 김 씨는 <선생님: CV_OCCUPATION>으로 일하고 있다.
 - 이 문장에서 '선생님'은 '~으로 일하고 있다'와 같이 공기하여 쓰였으므로 CV_OCCUPATION으로 태깅한다.
 - 김철수 선생님 되시지요? 여기 한라은행 춘천점입니다.
 - 우리 동네 의사 선생님은 참 친절하다.
 - '김철수 선생님', '직업명+선생님'에서 '선생님'은 '교사'가 아니라 남을 존대하는 말로 사용되었으므로 태깅 대상이 아니다.
- (태깅 대상 - 유튜버, 블로거) 유튜버는 최근 초등학생 장래희망직업 순위에도 오르기도 했다. 새로 생겨난 말 중에 직업으로 인정이 되는 말이라면 <우리말샘> 사전 등재 여부를 떠나 CV_OCCUPATION으로 태깅한다. 또한 직업 명칭이 구 구성으로 이루어진 경우 또한 사전 등재 여부와 관계없이 직업으로 인정된다면 최장 태깅 원칙을 적용하여 전체를 하나로 태깅한다.
 - 유튜버, 블로거 → <유튜버: CV_OCCUPATION>, <블로거: CV_OCCUPATION>
 - 파워 블로거 → <파워 블로거: CV_OCCUPATION>
- (태깅 대상 - 의병) '의병'은 군인의 일종이기 때문에 CV_OCCUPATION으로 태깅한다.
 - 의병 홍범도 장군 → <의병: CV_OCCUPATION> <홍범도: PS_NAME> <장군: CV_POSITION>
 - '장군'은 직위로 보고 CV_POSITION으로 태깅한다.
- (태깅 단위 - 운동 종목+선수) '운동 종목+선수'는 묶어서 CV_OCCUPATION으로 태깅한다. '축구 선수'가 우리말샘 검색 결과 '축구^선수'로 등재되어 있다. 마찬가지로 다른 운동 종목도 '선수'와 묶어서 태깅한다.
 - 축구 선수 → <축구 선수: CV_OCCUPATION>
- (태깅 범주 - 경찰) 경찰은 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.
 - 경찰에 신고하다. → <경찰: OGG_POLITICS>에 신고하다.
 - '교통'이라고 적힌 연두색 조끼를 입은 경찰을 배치했다. → '교통'이라고 적힌 연두색 조끼를 입은 <경찰: CV_OCCUPATION>을 배치했다.
 - 위 문장에서 '경찰'은 경찰 공무원을 의미하기 때문에 OGG_POLITICS로 태깅하지 않고 CV_OCCUPATION으로 태깅한다.

7.16. CV_POSITION

7.16.1. 정의 및 설명

- 공무원/군인/교수의 직위/직책/직급, 직장/학교/고대 국가에서의 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
- 군인의 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.

- 군인의 직위/직책/직급
 - 김 <병장: CV_POSITION>이 나를 쳐다 보았다.
- 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

- (태깅 대상 제외 - 양반, 노예) CV_POSITION에는 '직위/직책/직급'에 해당되는 것만 태깅한다. '양반, 노예, 수드라'와 같이 '직위/직책/직급'에 해당하지 않는 신분명은 태깅하지 않는다.
 - 양반, 노예, 수드라, 바이샤, 브라만, 크샤트리아
 - 태깅 대상 아님.
- (태깅 대상 제외 - 신부, 성직자) '신부, 성직자'는 CV_OCCUPATION으로 태깅되기 때문에 CV_POSITION으로 태깅하지 않는다.
- (태깅 단위 - 인명 + 지위, 지위 + 인명) 다음과 같은 명사구는 지위와 인명을 각각 따로 태깅한다.
 - 문재인 대통령 → <문재인: PS_NAME> <대통령: CV_POSITION>
 - 이재용 부회장 → <이재용: PS_NAME> <부회장: CV_POSITION>
 - 국대비 왕씨 → <국대비: CV_POSITION> <왕: PS_NAME> 씨
 - 폐비 윤씨 → <폐비: CV_POSITION> <윤: PS_NAME> 씨
 - 김 박사 → <김: PS_NAME> <박사: CV_POSITION>
 - 류샤오밍 대사 → <류샤오밍: PS_NAME> <대사: CV_POSITION>
- (태깅 단위 - 묘호+직위/직책+이름) 왕의 명칭의 경우, 묘호와 직위 및 직책을 분리하여 태깅한다. 단, 묘호나 시호 등이 직위와 함께 쓰여 한 단어로 쓰인 경우, 각각 따로 내부 태깅을 하지 않는다. 이는 한 사람을 지칭하는 말로 볼 수 있으므로 PS_NAME으로 태깅한다.
 - 조선 3대 태종왕 이방원의 묘 현릉에 다녀왔어요. → 조선 3대 <태종: PS_NAME><왕: CV_POSITION> 이방원의 묘 현릉에 다녀왔어요.
 - 정조 대왕 이산은 조선의 제22대 왕이다. → <정조: PS_NAME> <대왕: CV_POSITION> <이산: PS_NAME>은 조선의 제22대 왕이다.
 - 순정효황후 → <순정효황후: PS_NAME>

- 명성황후 → <명성황후: PS_NAME>
- 장수왕 → <장수왕: PS_NAME>
- 무열왕 → <무열왕: PS_NAME>
- (태깅 단위 - 지역명+직위) '지역명+직위'는 묶어서 CV_POSITION으로 태깅한다. '광주시장, 서울시장'이 우리말샘에 복합어로 등재되어 있다. 이와 같은 방식으로 시장, 도지사, 광역시장, 부시장, 군수, 부군수, 구청장 등의 지자체 공무원 직위/직책/직급의 경우, '지역명+직위' 전체를 CV_POSITION으로 태깅한다.
 - 광주 시장 → <광주 시장: CV_POSITION>, 사하구청장 → <사하구청장: CV_POSITION>
- (태깅 단위 - 기관+직위) '선거대책위원장'과 같이 접사 '-장'이 붙어서 나누어지지 않는 경우, 전체를 CV_POSITION으로 태깅한다. '환경부 장관, 남구청 부구청장'과 같이 나누어지는 경우는 우리말샘 검색 결과를 따른다. 검색이 되는 경우는 전체를 CV_POSITION으로 태깅, 검색이 되지 않는 경우는 OGG와 CV_POSITION으로 따로 태깅한다.
 - 선거대책위원장 → <선거대책위원장: CV_POSITION>
 - '선거대책위원장'은 '-장'이 포함된 복합어로 보고 분리하여 태깅하지 않는다.
 - 환경부 장관 → <환경부: OGG_POLITICS> <장관: CV_POSITION>
 - 남구청 부구청장 → <남구청: OGG_POLITICS> <부구청장: CV_POSITION>
 - '환경부 장관, 남구청 부구청장'은 우리말샘에서 검색되지 않으므로 따로 태깅한다.
- (태깅 단위 - 부서+직위) '환경정책과장, 언어정보과장' 등과 같이 접사 '-장'이 붙어서 나누어지지 않는 경우, 전체를 CV_POSITION으로 태깅한다. '삼성전자 상무'와 같이 나누어지는 경우는 우리말샘 검색 결과를 따른다. 검색이 되는 경우는 전체를 CV_POSITION으로 태깅, 검색이 되지 않는 경우는 직위만 CV_POSITION으로 따로 태깅한다.
 - 환경정책과장 → <환경정책과장: CV_POSITION>
 - 언어정보과장 → <언어정보과장: CV_POSITION>
 - '환경정책과장, 언어정보과장'은 '-장'이 포함된 복합어로 보고 분리하여 태깅하지 않는다.
 - 언어정보과 이승재 과장 → 언어정보과 이승재 <과장: CV_POSITION>
 - 부서와 직위가 분리되어 있고 '과장'은 '직장에서의 직위/직책/직급'에 해당하므로 '과장'만 CV_POSITION으로 태깅한다.
- (태깅 단위 - 전/현+직위) '현 사장, 전 국무총리'와 같이 직위에 '전, 현'이 포함되어 나올 경우 '전, 현'은 태깅하지 않는다.
 - 현 사장 → 현 <사장: CV_POSITION>

7.17. CV_FOOD

7.17.1. 정의 및 설명

- 요리의 유형, 요리명, 식품명, 식재료

7.17.2. 분석 대상

- 요리의 유형

- 국, 탕국, 밥, 반찬, 국수, 나물, 무침, 향신료, 안주, 술안주
- 요리명
 - 일반 요리명: 쇠고기 덮밥, 어리굴젓, 명란젓, 해물겨자채, 김치, 나박김치, 봄배추김치, 찐지, 오이소박이, 열무김치, 순살치킨, 양념치킨, 쌀밥, 보리밥
 - 식당, 음식점 등의 메뉴명: 붉은대게다리살죽, 어니언치킨, 파닭, 1955버거, 해피밀, 시간의 정성 진카스테라
- 식품명
 - 가공 식품명: 복어, 건태, 건명태, 건포도, 곱감, 라면 등
 - 가공 식품의 상품명: 진짬뽕, 신라면, 초코파이, 새우깡, 연두, 멸치디포리다시팩, 일판란, 냉동차돌박이바로구이, 새콤달콤꼬마유부초밥
- 식재료
 - 곡물의 명칭: 현미, 좁쌀, 울무, 백미
 - 음식 재료: 소고기, 돼지고기, 닭고기, 버섯, 열무, 파
 - 양념: 소금, 설탕, 간장, 겨자즙

7.17.3. 예문

- 요리의 유형
 - 한국에서도 아주 오래 전부터 <국수: CV_FOOD>를 만들어 먹었으나, 밀의 생산이 많지 않았기 때문에 상용음식이 되지는 못하였다.
 - <나물: CV_FOOD>은 그중 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> · <찐지: 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 맞추고 넣는 양이 적당해야 맛있다.

7.17.4. 주의 사항

- (태깅 대상 - 밥) '음식'의 의미인 '밥'과 '끼니, 식사'의 의미인 '밥' 모두 CV_FOOD로 태깅한다.
 - 이건 밥이랑 먹어야 맛있다. → 이건 <밥: CV_FOOD>이랑 먹어야 맛있다.
 - 음식의 의미인 '밥'으로 쓰였으므로 CV_FOOD로 태깅한다.

- 1시가 넘었는데 밥은 먹었어? → 1시가 넘었는데 <밥: CV_FOOD>은 먹었어?
 - 문장에서 '밥'은 '음식'이 아닌 '끼니, 식사'의 의미로 쓰였지만 CV_FOOD로 태깅한다.
- (태깅 단위 - 복합어) '초코 쿠키, 바닐라 아이스크림'과 같이 음식명에는 복합어로 이루어진 것이 많다. 음식명의 경우 복합어로 쓰인 것이 특정한 개체를 가리키므로 내부를 분리하지 않고 전체를 하나로 태깅한다.
 - <젤라또: CV_FOOD>부터 <장미아이스크림: CV_FOOD>까지 종류도 다양하고 맛도 좋은 아이스크림 맛집으로 초대합니다.
 - '젤라또'는 음식명이므로 CV_FOOD로 태깅한다. '장미아이스크림'은 '장미'와 '아이스크림'이 같이 쓰였을 때 특정한 개체를 가리키므로 '장미아이스크림' 전체를 태깅한다. <장미: PT_FLOWER><아이스크림: CV_FOOD>로 분리 태깅하지 않는다.
 - 쌀밥, 보리밥 → <쌀밥: CV_FOOD>, <보리밥: CV_FOOD>
- (태깅 단위 - 삶은 계란, 삶은 달걀) 관형사형 어미와 함께 쓰인 '삶은 계란, 삶은 달걀'의 경우는 우리 말샘 사전에 구 단위로 등재되어 있다. 따라서 관형사형 어미를 제외한 '계란, 달걀'만을 CV_FOOD로 태깅한다.
 - 삶은 계란 → 삶은 <계란: CV_FOOD>
- (태깅 범주 - CV_FOOD / AM계열 / PT계열) 음식 재료명으로 쓰였다고 판단되는 경우 AM 또는 PT에 속하는 개체라 할지라도 CV_FOOD로 태깅한다.
 - 저녁 반찬으로 <시금치: CV_FOOD>를 먹었다.
 - '시금치'를 음식으로 먹은 것이므로 CV_FOOD로 태깅한다.
 - <시금치: PT_GRASS>는 높이 30~60cm로 뿌리는 굵고 붉으며, 잎은 어긋나고 세모진 달걀 모양이다.
 - 이 문장에서 '시금치'는 음식을 가리키는 것이 아니라 식물 '시금치'를 칭하므로 PT_GRASS로 태깅한다.

7.18. CV_DRINK

7.18.1. 정의 및 설명

- 음료/주류의 하위 부류, 음료/주류의 상품명

7.18.2. 분석 대상

- 음료의 하위 부류
 - 주스, 오렌지주스, 토마토주스, 우유, 요구르트, 홍차, 에이드, 레모네이드, 콜라, 사이다, 물
- 주류의 하위 부류
 - 위스키, 칵테일, 샴페인, 소주, 맥주, 양주, 술
- 음료/주류의 상품명
 - 코카콜라, 칠성사이다, 카스, 테라
 - 카페 음료명: 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.
- 음료/주류의 상품명
 - <코카콜라: CV_DRINK>부터 얼음컵에 쿵쿵 따라 붓습니다.
 - 원래 술 마시는 거 안 좋아하는데 <테라: CV_DRINK> 왜 이리 맛있나요?
 - <앗!메리카노: CV_DRINK> 다 마셔갈 때 찰칵했네요.

7.18.4. 주의 사항

- (태깅 대상 - 물) 마시는 용도의 '물'은 음료의 하위 부류이다. 마시는 용도로 '물'이 사용된 경우, CV_DRINK로 태깅한다. 그 외 용도로 사용된 물은 태깅 대상이 아니다.
 - 저는 <커피: CV_DRINK> 괜찮습니다. <물: CV_DRINK> 주세요.
 - 문맥상 '마시는 물'을 가리키므로 물을 CV_DRINK로 태깅한다.
 - 수영을 배운 첫날이라서 물을 많이 먹었어.
 - 여기에서 '물'은 마시는 용도가 아니므로 태깅 대상이 아니다.
- (태깅 대상 제외 - 음료, 음료수) 주스, 우유, 콜라 등의 상위 개념인 '음료, 음료수'는 태깅 대상이 아니다.
 - 음료수 중에 콜라를 제일 좋아한다. → 음료수 중에 <콜라: CV_DRINK>를 제일 좋아한다.
 - 음료의 종류인 '콜라'는 CV_DRINK로 태깅하고 상위 개념인 '음료수'는 태깅하지 않는다.
- (태깅 단위 - OO주스) '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

- 음식 유형

7.19.2. 분석 대상

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

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

- (태깅 대상 제외 - 절식 등) '절식'은 음식 종류가 아니므로 CV_FOOD_STYLE로 태깅하지 않는다. 또한 '야식, 간식, 디저트, 스낵, 국물'은 특정한 개체를 가리키는 것이 아니므로 개체명으로 보기 어렵다. 따라서 이 명사들은 CV_FOOD_STYLE로 태깅하지 않는다.
 - 패스트푸드는 디저트지. → <패스트푸드: CV_FOOD_STYLE>는 디저트지.
 - 디저트는 개체명으로 볼 수 없기 때문에 태깅하지 않고, '패스트푸드'만 CV_FOOD_STYLE로 태깅한다.

7.20. CV_CLOTHING

7.20.1. 정의 및 설명

- 섬유/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, 면바지, 기모바지

- 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

- (태깅 대상 - 운동 의복) 수영복, 등산복 등의 운동 의복은 CV_SPORTS_INST로 태깅하지 않고 CV_CLOTHING으로 태깅한다.
 - 테니스 교실에 테니스채와 테니스복을 챙겨 갔다. → 테니스 교실에 <테니스 채: CV_SPORTS_INST>와 <테니스복: CV_CLOTHING>을 챙겨 갔다.
 - 테니스 도구인 '테니스 채'는 CV_SPORTS_INST로 태깅하고, '테니스복'은 옷의 종류로 CV_CLOTHING으로 태깅한다.
- (태깅 대상 제외 - 상위어) '바지, 모자, 상의, 하의, 양말, 가방, 신발' 등 상위어는 태깅하지 않는다.
 - 바지 중에서 카르고 팬츠가 가장 편하다. → 바지 중에서 <카르고 팬츠: CV_CLOTHING>가 가장 편

하다.

- 상위어인 '바지'는 태깅하지 않고 바지의 종류인 '카르고 팬츠'만 CV_CLOTHING으로 태깅한다.
- (태깅 대상 제외 - 장신구) '-을/를 입다, 신다, 쓰다', '-을 벗다'와 어울리지 않으면 태깅하지 않는다. 옷, 신발류, 양말류, 가방류는 이러한 서술어와 어울리지만 '귀걸이, 목걸이, 반지'와 같은 장신구류는 전혀 어울리지 않으므로 태깅하지 않는다.
 - 교복을 입을 때는 귀걸이를 하지 않는 편이다. → <교복: CV_CLOTHING>을 입을 때는 귀걸이를 하지 않는 편이다.
 - 의복의 유형인 '교복'은 CV_CLOTHING으로 태깅하지만 장신구류인 '귀걸이'는 태깅하지 않는다.
- (태깅 대상 제외 - 상품명) 의류, 스포츠용품 등 정확한 상품명에 쓰였을 경우에는 AFW_OTHER_PRODUCTS로 태깅한다.
 - 등산복 중에서는 아이더 '조셉 남성 아이스 쿨넥 폴로티'가 판매율 1위를 기록했다. → <등산복: CV_CLOTHING> 중에서는 아이더 '<조셉 남성 아이스 쿨넥 폴로티: AFW_OTHER_PRODUCTS>'가 판매율 1위를 기록했다.
- (태깅 단위 - 섬유 명칭+바지) '섬유 명칭+바지'의 경우, 우리말샘 사전에 참고하여 다음과 같이 태깅한다. 우리말샘에서 '-바지'를 검색하면 '면-바지', '기모^바지'로 등재되어 있다. '-', '^'의 표시가 있는 경우, 전체를 CV_CLOTHING으로 태깅한다. 한편, '마 바지', '가죽 바지'는 띄어쓰기가 된 구 구성이다. '마, 가죽'을 섬유 명칭으로 보고 '마, 가죽'만 분리하여 태깅한다.
 - 면바지 → <면바지: CV_CLOTHING>, 기모바지 → <기모바지: CV_CLOTHING>
 - 마 바지 → <마: CV_CLOTHING> 바지, 가죽 바지 → <가죽: CV_CLOTHING> 바지
 - 나일론 바지 → <나일론: CV_CLOTHING> 바지
 - '나일론 바지'의 경우, 우리말샘 사전에 등재되어 있지 않다. 이런 경우, '띄어쓰기'로 등재된 '마 바지', '가죽 바지'의 형태를 따라 '나일론'만 분리하여 태깅한다.

7.21. CV_BUILDING_TYPE

7.21.1. 정의 및 설명

- 건축양식 유형

7.21.2. 분석 대상

- 건축양식 유형
 - 기능주의건축, 노르만양식, 시토파건축, 에마누엘 양식, 그리스건축, 도리스양식
 - 유럽식 목조주택, 기와집, 널벽집, 한옥

7.21.3. 예문

- 현존하는 가장 오래되고 전형적인 <시토파건축: CV_BUILDING_TYPE>이다.
- 파타라가 <에마누엘 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 날짜 표현을 대상으로 한다.

분류 명칭	정의 및 설명
8.1. DT_DURATION	- 기간
8.2. DT_DAY	- 특정 날짜 및 절기 - 기념일(법정 기념일과 같이 특정 날짜를 알 수 있는 기념일) - 평일, 공휴일, 특정 날을 의미하는 단어
8.3. DT_WEEK	- 주(週)
8.4. DT_MONTH	- 월(月)
8.5. DT_YEAR	- 년(年)
8.6. DT_SEASON	- 계절
8.7. DT_GEOAGE	- 지질시대
8.8. DT_DYNASTY	- 왕조시대 - 역사적 시대
8.9. DT_OTHERS	- DT 계열의 세부 유형에 속하지 않는 기타 날짜 표현

- 주의 사항
 - (태깅 대상 - 보통명사) '오늘, 내일, 이달, 올해, 평일, 주말, 이번 주' 등과 같은 보통명사 또한 태깅한다. 각각에 해당하는 세부분류는 아래 지침에서 확인할 수 있다.
 - (태깅 대상 - 숫자만 쓰인 경우) 날짜 표현 없이 숫자 자체만으로도 문맥상 날짜를 의미하는 것이라면 DATE로 태깅한다.
 - 오늘은 2021/7/10입니다. → 오늘은 <2021/7/10: DT_OTHERS>입니다.
 - (태깅 대상 제외 - 매O) '매일, 매주, 매달, 매월, 매해, 매년'은 태깅하지 않는다.
 - (태깅 단위 - 연+월+일+요일 구성) '20/11/15'와 같은 구성은 통합하여 DT_OTHERS로 태깅한다. '년+월+일' 셋 중 둘만 함께 쓰이는 경우 및 영문도 포함한다. 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.
 - (태깅 단위 - 구체적 표현 포함) DATE와 함께 쓰여 의미를 구체화하는 '약~, ~여, ~쯤, ~께, 한~, ~정도, ~안팎, ~간, ~내내, ~동안, ~가량, ~만, ~씩, ~째, ~이내, ~안, ~부터 ~까지, ~에서 ~까지, ~전, ~중, ~후, ~때, 총~, 지난~, 오는~, ~초, ~말' 등과 같은 표현이 있다면 함께 태깅한다. 각각에 해당하는 세부분류는 아래 지침에서 확인할 수 있다.
 - (태깅 단위 - 지난/오는+DT) '지난/오는'은 단독 태깅하지 않으며, '지난+DT 범주에 속하는 명사 또는 명사구'일 경우에만 DT를 포함하여 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.

- 오는 4일에 행사를 개최할 예정이다. → <오는 4일: DT_DAY>에 행사를 개최할 예정이다.
 - '지난/오는 + DT'의 경우, '지난/오는'을 포함하여 태깅하며, 태깅 범주는 뒤에 오는 DT에 따른다.
- 손흥민은 지난 아시안게임에서 금메달을 땀다. → 손흥민은 지난 <아시안게임: EV_SPORTS>에서 금메달을 땀다.
 - '지난+EV'의 경우, '지난'을 제외하고 EV만 태깅한다.
- 손흥민은 지난 2014년 10월 아시안게임에서 금메달을 땀다. → 손흥민은 <지난 2014년 10월: DT_OTHERS> <아시안게임: EV_SPORTS>에서 금메달을 땀다.
 - '지난 2014년 10월'은 '지난+DT' 구성이므로 함께 태깅한다. EV인 '아시안게임'은 분리하여 태깅한다.
- (태깅 단위 - 날짜+순서+상) 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>
- (태깅 범주 - DT계열/QT계열) 'n일, n일째, n일 차'의 경우, 아래와 같이 태깅한다.
 - (n일) '얼마 동안'의 경우는 DT_DURATION, 달력상의 날짜는 DT_DAY, '일'을 단위로 날의 수를 세는 경우는 QT_COUNT로 태깅한다.
 - 커플 31일 기념 프사 → 커플 <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
 - 올해 공휴일은 31일밖에 안 된다. → 올해 공휴일은 <31일밖에 : QT_COUNT> 안 된다.
 - (n일째) '얼마 동안'의 의미로 쓰였으면 DT_DURATION, '순차'의 의미를 나타내면 QT_ORDER로 태깅한다.
 - 3일째 비가 내렸다. → <3일째: DT_DURATION> 비가 내렸다.
 - 3일째 되는 날 비가 내리기 시작했다. → <3일째: QT_ORDER> 되는 날 비가 내리기 시작했다.
 - (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 1일차 → 교생실습 <1일차: QT_ORDER>
 - 입사 8년차 → 입사 <8년차: QT_ORDER>

8.1. DT_DURATION

8.1.1. 정의 및 설명

- 기간

8.1.2. 분석 대상

- ~간, ~내내, ~동안, ~가량, ~만, ~씩
- ~부터 ~까지, ~에서 ~까지
- 전반기/후반기, 성수기/비수기, 환절기, 방학
- 하루, 이틀, 주말, 일주일, ~주, ~개월, ~한 해, ~년, N박 N일
- 월초, 월말, 연초, 연말
- ~시즌
- ~째

8.1.3. 예문

- ~간, ~내내, ~동안, ~가량, ~만, ~씩
 - 부민문화사에서 발행된 산림기사 기출문제 해설은 <10년간: DT_DURATION>의 기출문제를 모두 담 아 해설을 첨부하였습니다.
 - 안스리움은 <일년 내내: DT_DURATION> 피는 꽃이다.
 - 과거 <50년 동안: DT_DURATION> 한국이 가난에서 벗어나서 경제발전을 했던 그 경험들이 있잖아 요.
 - 가족을 <3일 만: DT_DURATION>에 만났다.
 - 방학 숙제를 <3일 안: DT_DURATION>에 해야 한다.
- ~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 어려 울 것 같네.
 - 철학의 원조라고 하는 고대 그리스의 철학자 탈레스(<BC624~BC546: DT_DURATION>)는 '짐승이 아 니라 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, 방학
 - <전반기: DT_DURATION> 근로장려금을 신청 못 했으면 <후반기: DT_DURATION>에 신청하면 됩니 다.
 - 제주도는 <성수기: DT_DURATION>에 가면 어딜 가도 사람이 많고 비싸니까, <비수기: DT_DURATION>에 가는 걸 추천해요.
 - <환절기: DT_DURATION> 건강 꿀조합템을 소개합니다.
 - 무순시 신화소학교는 <지난 겨울방학간: DT_DURATION> 학생수공작품 전람을 진행했다.
- 하루, 이틀, 주말, 일주일, ~주, ~개월, ~한 해, ~년, N박 N일
 - <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 떠보

니 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.
- 여름에 휴가는 가고 싶은 이들에게 <1박 2일: DT_DURATION>로 가볍게 다녀오기 좋은 속초 여행지를 소개하고자 한다.
- 월초, 월말, 연초, 연말
 - <연말: DT_DURATION>이면 설레고 희망적인 낙천적인 분위기가 떠오른다.
 - 벌써 <월말: DT_DURATION>, 그런데 월초<DT: DURATION>면 없다.
 - 마침 <연초: DT_DURATION>에 리프레시하기 좋은 겨울철 여행지를 추천해 드릴까 해요.
- ~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 짜: 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

- (태깅 대상 제외 - 방학) 보통명사로 쓰인 방학은 태깅하지 않는다. 특정한 기간을 뜻할 경우만 DT_DURATION으로 태깅한다.
 - 방학에는 특별 수업 때문에 학교에 나가야 한다. → 방학에는 특별 수업 때문에 학교에 나가야 한다.
 - '방학'이 특정한 기간을 뜻하지 않고 일반 명사로 쓰였기 때문에 태깅하지 않는다.
 - 무순시 신화소학교는 지난 겨울방학간, 학생수공작품 전람을 진행했다. → 무순시 신화소학교는 <지난 겨울방학간: DT_DURATION> 학생수공작품 전람을 진행했다.
 - '지난 겨울방학간'은 특정한 기간을 뜻하므로 DT_DURATION으로 태깅한다.
- (태깅 대상 제외 - D+n) 'D+6'과 같은 구성은 DT_DURATION이 아니라 QT_ORDER로 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 해석되기 때문에 QT_ORDER로 태깅한다.
- (태깅 대상 제외 - 평일) '평일'은 우리말샘 정의에 따르면 '토요일, 일요일, 공휴일이 아닌 보통 날'이며 '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.
 - 이곳은 평일에 안 막히는데 주말에만 막힌다. → 이곳은 <평일: DT_DAY>에 안 막히는데 <주말: DT_DURATION>에만 막힌다.

- (태깅 단위 - 구체성을 더하는 표현) DURATION과 함께 구체성을 더하는 표현(약~, -여, -쯤, 한~, ~정도, ~안팎, 총~)이 있을 경우 함께 태깅한다.
 - 이 작업은 한 4일 정도 걸린다. → 이 작업은 <한 4일 정도: DT_DURATION> 걸린다.
 - '4일'이라는 기간을 구체화하는 '한'과 '정도'도 같이 태깅한다.
- (태깅 단위 - 구체적인 DATE+초, 말) '초初, 말末'은 DATE의 구체성을 더해주므로 'DATE+초, 말' 전체를 태깅한다. '월초, 월말, 연초, 연말'과 같이 구체적인 DATE가 없이 '초, 말'과 결합한 경우에는 DT_DURATION으로 태깅한다. '7월 초, 2020년 말, 올해 초'와 같이 구체적인 DATE와 같이 쓰인 경우, 해당 DATE 범주로 태깅한다.
 - 월초, 월말, 연초, 연말 → <월초: DT_DURATION>, <월말: DT_DURATION>, <연초: DT_DURATION>, <연말: DT_DURATION>
 - 올해 초에 사업을 시작했다. → <올해 초: DT_YEAR>에 사업을 시작했다.
 - 6월 말에 학기가 끝난다. → <6월 말: DT_MONTH>에 학기가 끝난다.
 - '초, 말'이 '올해, 6월'이라는 구체적인 DATE와 같이 쓰였으므로 각각 DT_YEAR, DT_MONTH로 태깅한다.
- (태깅 단위 - 구체적인 DATE+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 않다면 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.
- (태깅 범주 - 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. 복합어 '하루하루' 또한 태깅하지 않는다. '하루하루'는 '그날그날의 날' 또는 '하루가 지날 때마다'의 뜻을 가지는데 DURATION이나 특정 DAY로 볼 수 없다. 또한 복합어의 내부 성분인 '하루'를 '이틀, 사흘' 등으로 교체할 수 없으므로 태깅하지 않는다.
 - 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.
 - 막연히 '어떤 날'을 지칭하므로 DT_DAY로 태깅한다.
 - 하루하루 그냥 보내고 있다. → 하루하루 그냥 보내고 있다.
 - '하루하루'는 태깅하지 않는다.
- (태깅 범주 - DT_DURATION/DT_WEEK) '~주'는 맥락에 따라 WEEK나 DURATION으로 태깅한다.
 - 7월 첫째 주 행사 → 7월 <첫째 주: DT_WEEK> 행사
 - 기간을 뜻하지 않기 때문에 DT_WEEK으로 태깅한다.
 - 특히 바지를 입었을 때 엉덩이가 터지기 직전이었었는데, 3주가 지난 지금은 푹 떨어지는 핏으로 바뀌었습니다. → 특히 바지를 입었을 때 엉덩이가 터지기 직전이었었는데, <3주: DT_DURATION>가 지난 지금은 푹 떨어지는 핏으로 바뀌었습니다.
 - 3주라는 기간을 뜻하므로 DT_DURATION으로 태깅한다.
- (태깅 범주 - N학기) 'N학기'가 '내내, 동안' 등과 함께 쓰여 지속성을 의미하는 것이라면

DT_DURATION, 학기를 세는 단위라면 QT_COUNT, 학기의 순서, 순차를 나타내면 QT_ORDER로 태깅한다.

- 2학기 내내 수학을 열심히 공부했다. → <2학기 내내: DT_DURATION> 수학을 열심히 공부했다.
- 지금까지 총 4학기를 들었다. → 지금까지 <총 4학기: QT_COUNT>를 들었다.
 - 지금까지 총 4번을 들은 것으로 해석된다. 세는 단위이기 때문에 QT_COUNT로 태깅한다.
- 2학기에 세부 전공을 정해야 한다. → <2학기: QT_ORDER>에 세부 전공을 정해야 한다.
 - 일반적인 두 번째 학기를 지칭하기 때문에 QT_ORDER로 태깅한다.
- (태깅 범주 - ~째) '~째'가 '동안'의 의미일 때만 DT_DURATION, '순서, 차례'의 의미일 경우에는 QT_ORDER로 태깅한다.
 - 3일째 비가 오고 있었다. → <3일째: DT_DURATION> 비가 오고 있었다.
 - 3일째 되던 날, 비가 그쳤다. → <3일째: QT_ORDER> 되던 날, 비가 그쳤다.
- (특수 항목 태깅 제외 - 다른 말 삽입) '~부터/에서 ~까지'의 경우 묶어서 DT_DURATION으로 태깅한다. '~부터/에서'와 '~까지' 중간에 다른 말이 삽입되어 있을 경우 묶어 태깅하지 않고 분리하여 태깅한다.
 - 6월부터 이게 8월까지 이어집니다. → <6월부터: DT_OTHERS> 이게 <8월까지: DT_OTHERS> 이어집니다.
 - 중간에 '이게'가 없다면 <6월부터 8월까지: DT_DURATION>으로 태깅해야 하지만 중간에 '이게'가 삽입되어 있을 경우 분리하여 각각 <6월부터: DT_OTHERS>, <8월까지: DT_OTHERS>로 태깅한다.

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, 특정 날을 의미하는 단어

8.2.2. 분석 대상

- N일
- 요일, 평일, 공휴일
- 절기
 - 입춘, 곡우, 하지 등
- 기념일 및 명절
 - 명절, 설날, 근로자의 날, 어버이날, 빼빼로데이, 크리스마스, 수능날, 정초 등
- 어제, 오늘, 금일, 당일, 내일, 모레, 첫날, 마지막날 등
- 특정 날을 의미하는 단어
 - 놀토, 불토

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 편한 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, 금일, 당일, 내일, 모레, 첫날, 마지막날 등
 - 국내 코로나19 확진자가 <어제: DT_DAY> 하루 30명 증가했습니다.
 - <오늘: DT_DAY> 오후면 서울 G20 정상회의는 마무리됩니다.
- 특정 날을 의미하는 단어
 - 무더운 여름 끝자락, 집에서도 <불금: DT_DAY> 분위기를 제대로 낼 수 있는 레드볼 칵테일 레시피 5가지를 소개합니다.

8.2.4. 주의 사항

- (태깅 대상 제외- 매일) '매일'은 태깅하지 않는다.
 - 그는 매일 공을 뒀다. → 그는 매일 공을 뒀다.
- (태깅 단위 - 구체적 표현 포함) DAY와 함께 구체성을 더하는 표현(지난~, 오는~, ~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 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.
- (태깅 단위 - N주 + 요일) '첫째 주 화요일'과 같은 구성이 있으면 각각 따로 태깅한다.
 - 셋째 주 수요일에 만날까요? → <셋째 주: DT_WEEK> <수요일: DT_DAY>에 만날까요?
- (태깅 단위 - n일 + 날) 'n일 + 날' 구성의 경우, 'n일'만으로도 DAY를 나타내지만 'n일 + 날' 전체를 DT_DAY로 태깅한다.
 - 10일 날 만나. → <10일 날: DT_DAY> 만나.
- (태깅 범주 - 하루) '하루'는 '막연히 지칭하는 어떤 날'을 의미하는 맥락이라면 DT_DAY로 태깅한다. '한 낮과 한 밤이 지나는 동안'을 의미하는 맥락이라면 DT_DURATION으로 태깅한다. 복합어 '하루하루'는 태깅하지 않는다. '하루하루'는 '그날그날의 날' 또는 '하루가 지날 때마다'의 뜻을 가지는데 DURATION이나 특정 DAY로 볼 수 없다. 또한 복합어의 내부 성분인 '하루'를 '이틀, 사흘' 등으로 교체할 수 없으므로 태깅하지 않는다.
 - 하루는 집에 와 보니 엄마가 없었다. → <하루: DT_DAY>는 집에 와 보니 엄마가 없었다.
 - 막연히 '어떤 날'을 지칭하므로 DT_DAY로 태깅한다.

- 시골 할머니 댁에서 하루를 보냈다. → 시골 할머니 댁에서 <하루: DT_DURATION>를 보냈다.
- 하루하루 그냥 보내고 있다. → 하루하루 그냥 보내고 있다.
 - '하루하루'는 태깅하지 않는다.

8.3. DT_WEEK

8.3.1. 정의 및 설명

- 주(週)

8.3.2. 분석 대상

- 지난주, 이번 주, 다음 주 등
- 첫 주, 둘째 주 등

8.3.3. 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.

8.3.4. 주의 사항

- (태깅 대상 제외 - 매주) '매주'는 태깅하지 않는다.
 - 매주 회의가 있다. → 매주 회의가 있다.
- (태깅 대상 제외 - 일주일) '일주일'은 월요일부터 일요일까지의 기간을 나타내는 것이므로 DT_WEEK가 아니라 DT_DURATION으로 태깅한다.
 - 일주일은 못 되더라도 2일은 머무를 각오들을 한다고 했다. → <일주일: DT_DURATION>은 못 되더라도 2일은 머무를 각오들을 한다고 했다.
- (태깅 단위 - 구체적 표현 포함) WEEK와 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.
 - 개강 후 둘째 주께 한 번 보자. → 개강 후 <둘째 주께: DT_WEEK> 한 번 보자.
- (태깅 단위 - N주+요일) '첫째 주 화요일'과 같은 구성이 있으면 각각 따로 태깅한다.
 - 셋째 주 수요일에 만날까요? → <셋째 주: DT_WEEK> <수요일: DT_DAY>에 만날까요?
- (태깅 범주 - DT_WEEK / DT_DURATION) '~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.
 - '4주간'이라는 기간을 뜻하므로 DT_DURATION으로 태깅한다.

- 시험 직전인 넷째 주에는 각 과목에 대한 복습과 취약점 보완을 중점으로 공부할 예정이야. → 시험 직전 <넷째 주: DT_WEEK>에는 각 과목에 대한 복습과 취약점 보완을 중점으로 공부할 예정이야.
 - '시험 직전 넷째 주'의 '넷째 주'는 기간이 아니라 특정한 주를 뜻하므로 DT_WEEK로 태깅한다.

8.4. DT_MONTH

8.4.1. 정의 및 설명

- 월(月)

8.4.2. 분석 대상

- N월
- 이달, 지난달, 다음 달 등
- 가정의 달, 호국보훈의 달 등 특정 달을 지칭하는 명칭

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
- 가정의 달, 호국보훈의 달 등 특정 달을 지칭하는 명칭
 - <가정의 달: DT_MONTH>을 맞아 선물 수요가 몰렸다.

8.4.4. 주의 사항

- (태깅 대상 제외 - 매월, 매달) '매월, 매달'은 태깅하지 않는다.
 - 매달 기부금을 고아원에 보내고 있다. → 매달 기부금을 고아원에 보내고 있다.
- (태깅 대상 제외 - ~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.
- (태깅 단위 - 구체적 표현 포함) MONTH와 함께 구체성을 더하는 표현(지난~, 오는~, 내달~, ~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 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.

- 춘천시는 시민 1인당 10만 원 가량의 재난지원금을 9월쯤 지급할 것이라 발표했다. → 춘천시는 시민 1인당 10만 원 가량의 재난지원금을 <9월쯤: DT_MONTH> 지급할 것이라 발표했다.
- (태깅 단위 - 구체적인 DATE + 초, 말) '초初, 말末'은 DATE의 구체성을 더해주므로 'DATE+초, 말' 전체를 태깅한다. '7월 초, 6월 말'와 같이 구체적인 MONTH와 같이 쓰인 경우 DT_MONTH로 태깅한다. DT_MONTH의 경우 중순(中旬)도 함께 태깅한다. '월초, 월말, 연초, 연말'과 같이 구체적인 DATE가 없이 '초, 말'과 결합한 경우에는 DT_DURATION으로 태깅한다.
 - 6월 말에 학기가 끝난다. → <6월 말: DT_MONTH>에 학기가 끝난다.
 - '말'이 '6월'이라는 구체적인 DATE와 같이 쓰였으므로 DT_MONTH로 태깅한다.
 - 독일 정부는 5월 중순에 분데스리가 무관중 리그를 재개하기로 하였다. → 독일 정부는 <5월 중순: DT_MONTH>에 분데스리가 무관중 리그를 재개하기로 하였다.
 - 월초, 월말, 연초, 연말 → <월초: DT_DURATION>, <월말: DT_DURATION>, <연초: DT_DURATION>, <연말: DT_DURATION>
- (태깅 단위 - n월+달) 'n월+달' 구성의 경우, 'n월'만으로도 MONTH를 나타내지만 'n월 + 달' 전체를 DT_MONTH로 태깅한다.
 - 9월 달은 바쁘다. → <9월 달: DT_MONTH>은 바쁘다.

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, 첫해, 마지막해 등
- ~년생
- 육십갑자+년
 - 을미년, 병자년 등
- 십이간지+해
 - 쥐띠 해 등
- 연호+연도
 - 영락 1년

8.5.3. 예문

- N년(대, 경 등)
 - 사피옌스에 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, 첫해, 마지막해 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.
- 연호+연도
 - <건양 1년: DT_YEAR>은 바로 태양력을 사용한 첫 해를 말한다.

8.5.4. 주의 사항

- (태깅 대상 제외 - 매년, 매해) '매년, 매해'은 태깅하지 않는다.
 - 매년 대학생이 줄고 있다. → 매년 대학생이 줄고 있다.
- (태깅 단위 - 연호+연도) '영락 1년'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
 - 건문은 중국 명나라 혜제 때의 연호다. → 건문은 중국 명나라 혜제 때의 연호다.
 - '영락 1년'은 '연호+연도' 구성이므로 DT_YEAR로 태깅하고, '건문'은 연호만 쓰였으므로 태깅하지 않는다.
- (태깅 단위 - 구체적 표현 포함) YEAR과 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.
 - 의정부시의 지식산업센터는 2022년쯤 완공될 예정이다. → 의정부시의 지식산업센터는 <2022년쯤: DT_YEAR> 완공될 예정이다.
- (태깅 단위 - 구체적인 DATE + 초, 말) '초初, 말末'은 DATE의 구체성을 더해주므로 'DATE+초, 말' 전체를 태깅한다. '2020년 초, 90년대 말'과 같이 구체적인 YEAR과 같이 쓰인 경우 DT_YEAR로 태깅한다. '월초, 월말, 연초, 연말'과 같이 구체적인 DATE가 없이 '초, 말'과 결합한 경우에는 DT_DURATION으로 태깅한다.
 - 박미경이라는 가수의 '민들레꽃씨 되어'는 90년대 초를 대표하는 유행곡입니다. → 박미경이라는 가수의 '민들레꽃씨 되어'는 <90년대 초: DT_YEAR>를 대표하는 유행곡입니다.
 - 월초, 월말, 연초, 연말 → <월초: DT_DURATION>, <월말: DT_DURATION>, <연초: DT_DURATION>, <연말: DT_DURATION>
- (태깅 단위 - 연도표시 + 스포츠 행사) 스포츠 행사가 연도와 병행하여 쓰일 경우 명확한 연도 표시는

DT로 태깅하고, 축약된 경우는 통합하여 EV로 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 타고르고 있다. → <88 서울올림픽경기: EV_SPORTS> 16일간의 대장정을 알리는 성화가 막 점화되어 타고르고 있다.

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) SEASON과 함께 구체성을 더하는 표현(올~, ~쯤, ~정도)이 있을 경우 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

- (태깅 단위 - 시대+OO기(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 역사적 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
 - 구한말, 일제강점기 등

8.8.3. 예문

- 왕조
 - 정약용은 <정조 시대: DT_DYNASTY> 청년기에 벼슬에 나가 정조의 깊은 총애를 받아 임금의 측근

에 있었으나 보수 집권 세력인 노론 벽파의 공격에 시달리며 순탄치 못한 관료기를 보내게 된다.

- 또한 <가리에누스 황제 시대: DT_DYNASTY>에 원로원을 군무에서 축출하는 법을 통과시키면서 군무와 정부 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.
 - <조선: DT_DYNASTY>에서 바로 일본 식민 지배로 넘어갔고 해방되자마자 소련 공산당의 영향권에 들어갔다.
 - 덕수궁은 <구한말: DT_DYNASTY> 어지러웠던 시기에 조선의 멸망을 지켜본 궁궐이다.

8.8.4. 주의 사항

- (태깅 대상 제외 - 현대) '현대'는 '고대'와 '중세'와 달리 시기를 확정할 수 없으므로 태깅하지 않는다.
 - 중세 시대 지어진 이 건물은 현대 건축물과 비교할 만하다. → <중세 시대: DT_DYNASTY> 지어진 이 건물은 현대 건축물과 비교할 만하다.
 - '중세 시대'만 DT_DYNASTY로 태깅하고 '현대'는 태깅하지 않는다.
- (태깅 단위 - 왕의 이름+때) 왕의 이름 뒤에 '~때'가 붙으면 DT_OTHERS가 아니라 DT_DYNASTY로 태깅한다. 이와 달리 특정 사건에 '~때'가 붙을 경우에는 DT_OTHERS로 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.
 - 이것은 한국전쟁 때 만들어졌다. → 이것은 <한국전쟁 때: DT_OTHERS> 만들어졌다.

8.9. DT_OTHERS

8.9.1. 정의 및 설명

- DT 계열의 세부 유형에 속하지 않는 기타 날짜 표현

8.9.2. 분석 대상

- 년+월+일, 년+월, 월+일
- ~ 전, ~후, ~뒤, ~중, ~내, ~안, ~이내 (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
- EV+때

8.9.3. 예문

- 년+월+일, 년+월, 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 공정률 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.
- EV+때
 - 당시 국군은 자체 전투복이 없어 <제2차 세계대전 때: DT_OTHERS> 미군이 입었던 군복 등을 입고 전쟁을 치렀다.

8.9.4. 주의 사항

- (태깅 대상 제외 - 왕의 이름 + 때) 왕의 이름 뒤에 '~때'가 붙으면 DT_OTHERS가 아니라 DT_DYNASTY로 태깅한다. 특정 사건에 '~때'가 붙을 경우에 DT_OTHERS로 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.
 - 이것은 한국전쟁 때 만들어졌다. → 이것은 <한국전쟁 때: DT_OTHERS> 만들어졌다.
- (태깅 단위 - 구체적 표현 포함) DATE와 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.

- 작년 1월경부터 한약을 먹기 시작했다. → <작년 1월경부터: DT_OTHERS> 한약을 먹기 시작했다.
- (태깅 단위 - EV+때) EV와 '때'가 같이 쓰였을 때 전체를 DT_OTHERS로 태깅한다.
 - 2002올림픽 때 재미있었다. → <2002올림픽 때: DT_OTHERS> 재미있었다.
 - 4라운드 때 상황이 반전되었다. → <4라운드 때: DT_OTHERS> 상황이 반전되었다.
- (태깅 단위 - 연/월/일+요일) '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.
- (태깅 단위 - 지난+DT/EV/OGG) '지난'은 단독 태깅하지 않으며, '지난+DT' 범주에 속하는 명사 또는 명사구일 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.
 - '아시안게임'은 DT 범주에 속하는 명사가 아니기 때문에 '지난'과 분리하여 태깅한다.
- (태깅 단위 - 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.
 - 손흥민은 지난 2014년 10월 아시안게임에서 금메달을 땀다. → 손흥민은 <지난 2014년 10월: DT_OTHERS> <아시안게임: EV_SPORTS>에서 금메달을 땀다.
 - '지난 2014년 10월'은 '지난+DT' 구성이므로 함께 태깅한다. 또한 '년+월' 구성이므로 DT_OTHERS로 태깅한다. '아시안게임'은 '때' 없이 쓰였으므로 DT가 아닌 EV로 태깅한다.
- (태깅 범주 - 연휴) '연휴'의 경우, 구체적인 날짜와 함께 쓰이지 않았다면 DT_OTHERS로 태깅한다.
 - 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.
 - 오는 5월 연휴에는 대만으로 여행을 가 보려고 해요. → <오는 5월 연휴: DT_DURATION>에는 대만으로 여행을 가 보려고 해요.

9. TIME

- TIME은 시계상으로 나타나는 시/시각이나 시간 범위를 대상으로 한다.

분류 명칭	정의 및 설명
9.1. TI_DURATION	- 기간
9.2. TI_HOUR	- 시간 및 시각 : 시간의 어느 한 시점, 하루의 24분의 1이 되는 동안을 세는 단위
9.3. TI_MINUTE	- 분, 반 : 한 시간의 60분의 1이 되는 동안을 세는 단위
9.4. TI_SECOND	- 초 : 한 시간의 3,600분의 1이 되는 동안을 세는 단위
9.5. TI_OTHERS	- TI 계열의 세부 유형에 속하지 않는 기타 시간

- 주의 사항
 - (태깅 대상 - 보통명사) '오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤'과 같은 보통명사 또한 태깅한다. 각각에 해당하는 세부분류는 아래 지침에서 확인할 수 있다.
 - (태깅 대상 - 숫자만 쓰인 경우) 시간 표현 없이 숫자 자체만으로도 문맥상 시간을 의미하는 것이라면 TIME으로 태깅한다.
 - 영업 시간: 매일 <10:30 ~ 21:00: TI_DURATION>
 - (태깅 단위 - 시+분+초) '시+분+초, 시+분, 분+초'가 함께 쓰일 경우에는 전체를 묶어서 태깅한다. 스포츠의 기록인 '02:07:20, 03'15", 05:37'과 같은 구성은 통합하여 TI_OTHERS로 태깅한다.
 - 현재 시각은 <11시 59분 57초: TI_OTHERS>입니다.
 - <12:00: TI_OTHERS>에 출발
 - 한국 남자 마라톤 신기록은 이봉주 선수가 세운 <02:07:20: TI_OTHERS>이다.
 - (태깅 단위 - 보통명사+시간) '오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤, am, pm, GMT'와 같은 보통명사가 단독으로 나올 경우 TI_DURATION으로 태깅한다. 뒤에 다른 시간 표현과 같이 나올 경우 묶어서 태깅하며, 범주는 뒤에 나온 시간 표현에 맞춘다.
 - 나른한 봄날 <오후: TI_DURATION>
 - 그 백화점의 개점 시간은 오전 10시이다. → 그 백화점의 개점 시간은 <오전 10시: TI_HOUR>이다.
 - (태깅 단위 - 구체적 표현 포함) TIME과 함께 쓰여 의미를 구체화하는 표현이 있다면 함께 태깅한다. 각각에 해당하는 세부 분류는 아래 지침에서 확인할 수 있다.
 - (태깅 범주 - TI계열) 특정 시각만 TI_HOUR, TI_MINUTE, TI_SECOND에 해당한다. 지속되는 시간은 TI_DURATION으로 태깅하고 '특정 시각+전, 후, 부터, 까지'를 나타낼 때는 TI_OTHERS로 태깅한다.
 - 1시에 식당에 도착해서 1시간 만에 다 먹었다. → <1시: TI_HOUR>에 식당에 도착해서 <1시간 만: TI_DURATION>에 다 먹었다.
 - 특정 시각인 '1시'는 TI_HOUR, 지속되는 기간을 뜻하는 '1시간 만'은 TI_DURATION으로 태깅한다.

- 1시부터 운동을 했다. → <1시부터: TI_OTHERS> 운동을 했다.
 - '1시부터'는 '시각+부터' 구성이므로 TI_OTHERS로 태깅한다.
- 영화 시작 <30분 전: TI_OTHERS>에 도착했어요.
 - '30분 전'은 '시간+전'의 구성으로 TI_OTHERS로 태깅한다.

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
 - ~간, ~내내, ~동안, ~가량, ~만, ~씩, ~째
 - 오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤
- 어떤 시각에서 어떤 시각까지의 사이
 - ~부터 ~까지

9.1.3. 예문

- 지속되는 기간
 - 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.
 - 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 올때 따뜻한 우유나 차를 마시고 자면 꼭 잘 수 있다.
- 어떤 시각에서 어떤 시각까지의 사이

- <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

- (태깅 대상 제외 - 복합어) '낮잠', '어젯밤', '여름밤'과 같은 복합어는 내부 태깅하지 않는다. 즉, '낮잠, 어젯밤, 여름밤' 모두 태깅하지 않는다.
 - 낮잠, 어젯밤, 여름밤 → 낮잠, 어젯밤, 여름밤
- (태깅 대상 제외 - 아침, 점심, 저녁) 끼니로 먹는 음식을 나타내는 '아침, 점심, 저녁'은 태깅 대상이 아니다.
 - 아침에 밥을 먹었다. → <아침: TI_DURATION>에 밥을 먹었다.
 - 이 문장에서 '아침'은 시간대를 의미하므로 TI_DURATION으로 태깅한다.
 - 오늘 아침을 못 먹었다. → 오늘 아침을 못 먹었다.
 - 이 문장에서 '아침'은 '아침밥'을 의미하므로 태깅하지 않는다.
- (태깅 대상 제외 - 밤낮없이) 부사 '밤낮없이'는 태깅 대상이 아니다.
 - 이틀동안 밤낮없이 비가 내렸다. → 이틀동안 밤낮없이 비가 내렸다.
- (태깅 단위 - OO 연속) '3시간 연속'에서 '연속'은 태깅하지 않는다.
 - 3시간 연속 → <3시간: TI_DURATION> 연속
 - '연속'은 태깅하지 않고 '3시간'만 TI_DURATION으로 태깅한다.
- (태깅 단위 - 구체적 표현 포함) DURATION과 함께 구체성을 더하는 표현(약~, ~여, ~쯤, 한~, ~정도, ~안팎, 총~)이 있을 경우 함께 태깅한다.
 - 주차하는 데 10여 분 정도 걸렸다. → 주차하는 데 <10여 분 정도: TI_DURATION> 걸렸다.
- (태깅 범주 - ~째) '~째'가 '동안'의 의미일 때만 TI_DURATION, '순서, 차례'의 의미일 경우에는 QT_ORDER로 태깅한다.
 - 1시간째 비가 오고 있었다. → <1시간째: TI_DURATION> 비가 오고 있었다.
- (특수 항목 태깅 - 기호 삽입) 시간을 나타내는 개체명 내부의 기호(+, -, x, :)는 포함하여 태깅한다. 수학적 기호(+, -, ÷, ×, /)가 포함된 경우 전체를 QT_OTHERS로 태깅한다.
 - 2+2 총 4시간 → <2+2: QT_OTHERS> <총 4시간: TI_DURATION>
- (특수 항목 태깅 예외 - 다른 말 삽입) '~부터/에서 ~까지'의 경우 묶어서 TI_DURATION으로 태깅한다. '~부터/에서'와 '~까지' 중간에 다른 말이 삽입되어 있을 경우 묶어 태깅하지 않고 분리하여 태깅한다.
 - 9시부터 이게 10시까지 이어집니다. → <9시부터: TI_OTHERS> 이게 <10시까지: TI_OTHERS> 이어집니다.
 - 중간에 '이게'가 없다면 <9시부터 10시까지: TI_DURATION>으로 태깅해야 하지만 중간에 '이게'가 삽입되어 있을 경우 분리하여 각각 <9시부터: TI_OTHERS>, <10시까지: TI_OTHERS>로 태깅한다.

9.2. TI_HOUR

9.2.1. 정의 및 설명

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

9.2.2. 분석 대상

- 시간 및 시각
 - 시
 - 오전, 오후, 아침, 점심, 저녁, 새벽, 낮, 밤 + 시

9.2.3. 예문

- 시간 및 시각
 - 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

- (태깅 단위 - 구체적 표현 포함) HOUR와 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.
 - 한 2시쯤 친구를 만났다. → <한 2시쯤: TI_HOUR> 친구를 만났다.
- (태깅 범주 - TI_HOUR / TI_DURATION / TI_OTHERS) 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.
 - 1시에 식당에 도착해서 1시간 만에 다 먹었다. → <1시: TI_HOUR>에 식당에 도착해서 <1시간 만: TI_DURATION>에 다 먹었다.
 - 특정 시각인 '1시'는 TI_HOUR, 지속되는 기간을 뜻하는 '1시간 만'은 TI_DURATION으로 태깅한다.
 - 1시부터 운동을 했다. → <1시부터: TI_OTHERS> 운동을 했다.
 - '1시부터'는 '시각+부터' 구성이므로 TI_OTHERS로 태깅한다.

9.3. TI_MINUTE

9.3.1. 정의 및 설명

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

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

- (태깅 단위 - 구체적 표현 포함) MINUTE와 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.
 - 아까 한 30분께 나왔다. → 아까 <한 30분께: TI_MINUTE> 나왔다.
- (태깅 범주 - TI_MINUTE / TI_DURATION / TI_OTHERS) 특정 분만 TI_MINUTE에 해당한다. 지속되는 시간은 TI_DURATION으로 태깅하고 '특정 분+전, 후, 부터, 까지'를 나타낼 때는 TI_OTHERS로 태깅한다. 특정 분만 해당하는 TI_MINUTE의 태깅과 구별하여야 한다.
 - 30분에 사진이 나왔는데 고르는 데만 40분 걸렸다. → <30분: TI_MINUTE>에 사진이 나왔는데 고르는 데만 <40분: TI_DURATION> 걸렸다.
 - 특정 분인 '30분'은 TI_MINUTE, 지속되는 기간을 뜻하는 '40분'은 TI_DURATION으로 태깅한다.
 - 30분부터 운동을 했다. → <30분부터: TI_OTHERS> 운동을 했다.
 - '30분부터'는 '분+부터' 구성이므로 TI_OTHERS로 태깅한다.

9.4. TI_SECOND

9.4.1. 정의 및 설명

- 초
 - 한 시간의 3,600분의 1이 되는 동안을 세는 단위

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

- (태깅 단위 - 구체적 표현 포함) SECOND와 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~

안팎)이 있을 경우 함께 태깅한다.

- 친구는 12시 59분 50초에 도착했고, 나는 55초경에 도착했다. → 친구는 12시 59분 50초에 도착했고, 나는 <55초경: TI_SECOND>에 도착했다.
- (태깅 범주 - TI_SECOND / TI_DURATION / TI_OTHERS) 특정 초만 TI_SECOND에 해당한다. 지속되는 시간은 TI_DURATION으로 태깅하고 '특정 초+전, 후, 부터, 까지'를 나타낼 때는 TI_OTHERS로 태깅한다. 특정 초만 해당하는 TI_SECOND의 태깅과 구별하여야 한다.
 - 57초에 결승선에 들어왔다. → <57초: TI_SECOND>에 결승선에 들어왔다.
 - 진동이 3초 정도 계속되었다. → 진동이 <3초 정도: TI_DURATION> 계속되었다.
 - 특정 초인 '57초'는 TI_SECOND, 지속되는 기간을 뜻하는 '3초'는 TI_DURATION으로 태깅한다.
 - 59초까지 제출이 가능하다. → <59초까지: TI_OTHERS> 제출이 가능하다.
 - '59초까지'는 '초+까지' 구성이므로 TI_OTHERS로 태깅한다.

9.5. TI_OTHERS

9.5.1. 정의 및 설명

- TIME 계열의 세부 유형에 속하지 않는 기타 시간

9.5.2. 분석 대상

- 시+분+초, 시+분, 분+초
- 00:00:00
- 시간/특정 시각(시, 분, 초)+~전, ~후, ~이전, ~이후, ~부터, ~까지, ~이내, ~안
('~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.
 - 망키님과 <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

- (태깅 대상 - 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_DURATION이 아닌 TI_OTHERS로 태깅한다.
 - 한국 남자 마라톤 신기록은 이봉주 선수가 세운 <02:07:20: TI_OTHERS>이다.
- (태깅 단위 - 구체적 표현 포함) TIME과 함께 구체성을 더하는 표현(~경, ~쯤, ~께, 한~, ~정도, ~안팎)이 있을 경우 함께 태깅한다.
 - 오전 10시 30분쯤에 지하철이 도착했다. → <오전 10시 30분쯤: TI_OTHERS>에 지하철이 도착했다.

10. QUANTITY

- QUANTITY는 수량, 분량이나 순서, 순차 표현을 대상으로 한다.

분류 명칭	정의 및 설명
10.1. QT_AGE	- 사람이나 동·식물 등이 태어나 살아온 햇수(연령), 개월 수(월령), 주수[(생후) n주], 일수[(생후) n일]
10.2. QT_SIZE	- 넓이, 면적 : 일정한 평면에 걸쳐 있는 공간이나 범위의 크기
10.3. QT_LENGTH	- 길이, 거리, 높이, 두께 : 한쪽 끝에서 다른 쪽 끝까지의 거리
10.4. QT_COUNT	- 수량, 빈도 : 사물의 수량이나 빈도 등을 세는 표현
10.5. QT_MAN_COUNT	- 인원수 : 사람의 수를 세는 표현
10.6. QT_WEIGHT	- 무게 : 사물의 무거운 정도를 나타내는 표현
10.7. QT_PERCENTAGE	- 백분율, 비율, 농도 : 비율과 관련된 표현
10.8. QT_SPEED	- 속도
10.9. QT_TEMPERATURE	- 온도, 기온, 수온 : 뜨거움과 차가움의 정도를 나타내는 표현
10.10. QT_VOLUME	- 부피, 용량, 용적 : 물체나 물질이 공간에서 차지하는 크기
10.11. QT_ORDER	- 순서, 순차, 등급
10.12. QT_PRICE	- 값, 가격, 금액 : 어떤 대상에 일정하게 매겨진 액수
10.13. QT_PHONE	- 전화번호
10.14. QT_SPORTS	- 스포츠 관련 수량 표현
10.15. QT_CHANNEL	- 텔레비전·라디오·무선 통신 따위의 주파수, 채널 번호
10.16. QT_ALBUM	- 노래 또는 연주곡 따위의 음반과 관련된 순차 표현
10.17. QT_ADDRESS	- 주소 관련 숫자 표현
10.18. QT_OTHERS	- QT 계열의 세부 유형에 속하지 않는 기타 수량 표현

- 주의 사항
 - (태깅 대상 - 환산 수치) 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>)
 - 1mile(1,760yd) → <1mile: QT_LENGTH>(<1,760yd: QT_LENGTH>)
 - (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우 괄호 안의 단위는 태깅하지 않는다.
 - 1제곱미터(m²) → <1제곱미터: QT_SIZE>(m²)
 - 1경(頃) → <1경: QT_SIZE>(頃)

- 2온스(oz) → <2온스: QT_WEIGHT>(oz)
- (태깅 단위 - 구체적 표현 포함) QUANTITY와 함께 쓰여 의미를 구체화하는 표현이 있다면 함께 태깅한다. 각각에 해당하는 세부 분류는 아래 지침에서 확인할 수 있다.
 - 약~, ~여, ~쯤, 한~, ~정도, ~안팎, ~가량, ~씩, ~째, ~만, ~이내, ~부터 ~까지, 총~, ~이상, ~이하, ~밖에
- (태깅 단위 - 범위로 나타난 경우) 'n1부터 n2까지', 'n1~n2'와 같이 범위를 나타낼 경우 'n1'의 분류 기준에 따라 전체를 태깅한다.
 - pp. 55~67 → <pp. 55~67: QT_ORDER>
 - 4세부터 7세까지 → <4세부터 7세까지: QT_AGE>
- (태깅 범주 - 단위 표지) SIZE / LENGTH / WEIGHT / PERCENTAGE / SPEED / TEMPERATURE / VOLUME은 실제로 사용되는 단위 표지를 기준으로 태깅한다. 각 항목의 "분석 대상"을 참고하여 태깅한다.
 - (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)
- (태깅 범주 - QT계열 / DT계열) 'n일, n일째, n일 차'의 경우, 아래와 같이 태깅한다.
 - (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 <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.
 - 3일 5시 출발 → <3일: DT_DAY> 5시 출발
 - (n일째) '순차'의 의미를 나타내면 QT_ORDER, '얼마 동안'의 의미로 쓰였으면 DT_DURATION으로 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>
- (특수 항목 태깅 예외 - 다른 말 삽입) QT의 경우 '숫자+단위'를 함께 태깅한다. 중간에 다른 말이 삽입되어 있을 경우 묶어 태깅하지 않고 분리하여 태깅한다.
 - 평균온도가 영하 이제 이백 도를 넘어갑니다. → 평균온도가 영하 이제 <이백 도: QT_TEMPERATURE>를 넘어갑니다.
 - <영하 이백 도: QT_TEMPERATURE> 전체 태깅인데 중간에 '이제'가 삽입되어 있으므로 '이백

도'만 태깅한다.

10.1. QT_AGE

10.1.1. 정의 및 설명

- 사람이나 동·식물 등이 태어나 살아온 햇수(연령), 개월 수(월령), 주수[(생후) n주], 일수[(생후) n일]

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 전사회는 <14세 이상부터: QT_AGE> 관람 가능합니다.
- 동물의 나이
 - 우리집 고양이 다다는 <다섯 살: QT_AGE>, 호야는 <네 살: QT_AGE>인데, 사람 나이로 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

- (태깅 대상 - 숫자만 나온 경우) '-살', '-세' 등의 단위가 없어도 맥락상 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.
- (태깅 대상 - n일/주) 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
- (태깅 대상 제외 - 백일(百日)) 일령과 관련하여, 백일(百日)이 행사의 의미를 가질 때에는 QT로 태깅하지 않는다.
 - (생후) 100일 된 아기 → (생후) <100일: QT_AGE> 된 아기
 - 시환이 백일에 와 주셔서 감사합니다. → 태깅하지 않음.
 - 이 문장에서 '백일'은 '백일 잔치'를 의미하므로 태깅하지 않는다.
- (태깅 범주 - D+n) 'D+n'이 '생후 n일'을 의미할 때에는 QT_AGE로 태깅하고 순차를 의미할 경우에는 QT_ORDER로 태깅한다.
 - 빈나네 마당 냥이 4남매 D+100!! 백일잔치~츄르잔치~ → 빈나네 마당 냥이 4남매 <D+100: QT_AGE> 백일잔치~츄르잔치~
 - 문맥상 생후 n일째의 의미이므로 QT_AGE로 태깅한다.
 - 초등학생 1학년 D+6, 오늘은 남산 나들이를 했습니다. → 초등학생 1학년 <D+6: QT_ORDER>, 오늘은 남산 나들이를 했습니다.
 - '생후'의 의미가 아닌 초등학생이 된 지 여섯 번째 날에 남산 나들이를 한 것이므로 순차로 보고 QT_AGE가 아닌 QT_ORDER로 태깅한다.

10.2. QT_SIZE

10.2.1. 정의 및 설명

- 넓이, 면적 : 일정한 평면에 걸쳐 있는 공간이나 범위의 크기

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 <1제곱미터: 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

- (태깅 대상 - 환산 수치) 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>)
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 1제곱미터(m²) → <1제곱미터: QT_SIZE>(m²)
 - 1경(頃) → <1경: QT_SIZE>(頃)

10.3. QT_LENGTH

10.3.1. 정의 및 설명

- 길이, 거리, 높이, 두께 : 한쪽 끝에서 다른 쪽 끝까지의 거리

10.3.2. 분석 대상

- 길이, 거리, 높이, 두께
 - 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, 1억 5000만 광년

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.
 - <약 1억5000만 광년: QT_LENGTH> 거리에 있는 거대한 블랙홀 하나
- 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.
- 두께
 - 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.

10.3.4. 주의 사항

- (태깅 대상 - 환산 수치) 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
 - 1mile(1,760yd) → <1mile: QT_LENGTH>(<1,760yd: QT_LENGTH>)
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 2미터(m) → <2미터: QT_LENGTH>(m)
- (태깅 대상 제외 - 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 항공캔넬은 가로 350mm, 세로 500mm, 높이 650mm입니다. → 펫모드의 항공캔넬은 가로 <350mm: QT_LENGTH>, 세로 <500mm: QT_LENGTH>, 높이 <650mm: QT_LENGTH>입니다.

10.4. QT_COUNT

10.4.1. 정의 및 설명

- 수량, 빈도 : 사물의 수량이나 빈도 등을 세는 표현

10.4.2. 분석 대상

- 수량 (n+[사람]을 제외한 보통명사)
 - 6개, 8백 70선, 60개국, 1마리, 두 마리, 약 600종, 약 3만 종, 약 1,000종, 9~10마리, 14종, 10권, 9개, 7가지, 여섯 가지, 5,000자, 4,123석, 390여 칸, 18행, 13척, 1,600여 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 (n+[사람]을 제외한 보통명사)
 - 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!!
 - <질량수 79: QT_COUNT> - 존재비율 50% <질량수 81: QT_COUNT> - 존재비율 50%
 - 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.
- 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 소리

가 마치 단단한 물질처럼 우리를 뺨 둘러쌌다.

- <세 겹: QT_COUNT>으로 쓴 꾸러미.
- D-n(일)
 - 강원 평창군 용평리조트와 발왕산 일대에서 개최된 2021 평창평화포럼 <D-100일: QT_COUNT> 붐업 평화 릴레이 걷기행사가 3일 성황리에 마무리됐다.
 - 수능 <D-3: QT_COUNT>!!!

10.4.4. 주의 사항

- (태깅 대상 - 한번, 한잔) 부사 '한번'은 태깅 대상이 아니다. (띄어 쓰지 않음/ "두 번", "세 번" 등으로 교체해서 어색하면 부사) 마찬가지로 '간단하게 한 차례 마시는 차나 술 따위'라는 뜻의 '한잔'도 태깅하지 않는다.
 - 언제 밥 한번 먹자.
 - 제가 한번 해볼게요.
 - 한번은 이런 일도 있었다니까요.
 - 그 녀석 목소리 한번 크네.
 - 위 네 문장의 '한번'은 '두 번, 세 번'으로 교체할 때 어색하다. 부사 '한번'으로 쓰였으므로 태깅하지 않는다.
 - 감식초 두 잔 할래, 까나리 한 잔 할래? → 감식초 <두 잔: QT_COUNT> 할래, 까나리 <한 잔: QT_COUNT> 할래?
 - 이 문장에서 '한 잔'은 '두 잔, 세 잔'으로 교체해도 어색하지 않으므로 QT_COUNT로 태깅한다.
- (태깅 대상 - n관왕) 'n관왕'은 특정 스포츠 관련 용어가 아니고 '아카데미 3관왕'과 같이 다른 분야에서 사용할 수 있기 때문에 QT_COUNT로 태깅한다.
 - 수원시청은 김나리-홍승연이 복식에도 결승에 올라, 단체전 우승에 이어 3관왕을 바라보게 됐다. → 수원시청은 김나리-홍승연이 복식에도 결승에 올라, 단체전 우승에 이어 <3관왕: QT_COUNT>을 바라보게 됐다.
- (태깅 대상 제외 - 오륜, 십계명) 오륜, 십계명과 같은 경우는 태깅하지 않는다.
- (태깅 범주 - 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 <31일밖에: QT_COUNT> 안 된다.
 - 이 문장에서 '31일'은 공휴일의 수를 센 것이므로 QT_COUNT로 태깅한다.
 - 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
 - 이 문장에서 '31일'은 특정 날을 의미하므로 DT_DAY로 태깅한다.
- (태깅 범주 - n회) 횟수를 의미하는 n회는 QT_COUNT로 태깅하고, 회차를 의미하는 n회는 QT_ORDER로 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”
- (태깅 범주 - n권) 책의 권수는 QT_COUNT로 태깅하고, 발행 순서상 '권'은 QT_ORDER로 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.
- (태깅 범주 - N학기) 'N학기'가 학기를 세는 단위라면 QT_COUNT, 학기의 순서, 순차를 나타내면 QT_ORDER, '내내, 동안' 등과 함께 쓰여 지속성을 의미하는 것이라면 DT_DURATION으로 태깅한다.
 - 지금까지 총 4학기를 들었다. → 지금까지 <총 4학기: QT_COUNT>를 들었다.
 - 지금까지 총 4번을 들은 것으로 해석된다. 세는 단위이기 때문에 QT_COUNT로 태깅한다.
 - 2학기에 세부 전공을 정해야 한다. → <2학기: QT_ORDER>에 세부 전공을 정해야 한다.
 - 일반적인 두 번째 학기를 지칭하기 때문에 QT_ORDER로 태깅한다.
 - 2학기 내내 수학을 열심히 공부했다. → <2학기: DT_DURATION> 내내 수학을 열심히 공부했다.

10.5. QT_MAN_COUNT

10.5.1. 정의 및 설명

- 인원수 : 사람의 수를 세는 표현

10.5.2. 분석 대상

- 인원수: n+명, 인, 분 등
 - 50명, 1명, 약 600명, 12인, 2억 1,219만 5,000명, 한 명 등
- n+[사람]에 속하는 보통명사
 - 대여섯 사람, 다섯 타자, 세 학생, 두 남자, 한 꼬마, 두 정상 등
- n+인승, 인용, 인분, 인방 등
 - 4인승, 2인용, 3인분, 3인방, 오인방 등

10.5.3. 예문

- 인원수: n+명, 인, 분 등
 - 경찰은 주지훈씨 외에도 <10여 명: QT_MAN_COUNT>의 마약투약 혐의를 포착하고 혐의를 입증하는데 주력하고 있어 연예계 마약파문은 당분간 계속 이어질 것 같습니다.
 - 많은 스타프들에게 둘러싸여 충실력을 뽐내고 있는 <여섯 명: QT_MAN_COUNT>의 스타들.
 - 이 쪽에 <한두 명만: QT_MAN_COUNT> 더 와 봐.
- n+[사람]에 속하는 보통명사
 - 지나가던 <세 학생: QT_MAN_COUNT>의 도움으로 위기를 모면할 수 있었다.

- 편의점 앞에서 <두 남자: QT_MAN_COUNT>가 싸우고 있었다.
- <한 꼬마: QT_MAN_COUNT>, <두 꼬마: QT_MAN_COUNT>, <세 꼬마: QT_MAN_COUNT> 인디언
- <세 타자: QT_MAN_COUNT> 연속 삼진
- 한일 <두 정상: QT_MAN_COUNT> 간의 만남에 세계의 이목이 집중되고 있다.
- n+인승, 인용, 인분, 인방 등
 - 다음은 덕촌리 꽃할매 <오인방: QT_MAN_COUNT>입니다.
 - 떡볶이 <5인분: QT_MAN_COUNT>을 셋이서 먹었다.
 - 현재 롤스로이스가 판매하고 있는 컨버터블 '던'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

- (태깅 대상 - 숫자만 나온 경우) 단위가 붙어있지 않아도 맥락상 사람 수를 의미하면 QT_MAN_COUNT로 태깅한다.
 - 과학자 둘이 다같이 한 거라고 주장하고 있다. → 과학자 <둘: QT_MAN_COUNT>이 다같이 한 거라고 주장하고 있다.

10.6. QT_WEIGHT

10.6.1. 정의 및 설명

- 무게 : 사물의 무거운 정도를 나타내는 표현

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, 30kg, 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 <5톤: QT_WEIGHT>이래요.

10.6.4. 주의 사항

- (태깅 대상 - 환산 수치) 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
 - 100냥(3,750g) → <100냥: QT_WEIGHT>(<3,750g: QT_WEIGHT>)

- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 2온스(oz) → <2온스: QT_WEIGHT>(oz)

10.7. QT_PERCENTAGE

10.7.1. 정의 및 설명

- 백분율, 비율, 농도 : 비율과 관련된 표현

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, 반, 절반 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%, 196proof, 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 줄어들죠.
 - 밥을 <반: QT_PERCENTAGE>만 먹었다
 - 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

- (태깅 대상 - 환산 수치) 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>)
- (태깅 대상 - 반, 절반) QT는 '숫자+단위'가 같이 나올 경우만 전체를 태깅하지만 '반, 절반'의 경우 50%를 의미하므로 예외적으로 QT_PERCENTAGE로 태깅한다.
 - 밥을 <반: QT_PERCENTAGE>만 먹었다.
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 40퍼센트(%) → <40퍼센트: 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 100km까지: 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 연료병으로 뚫고 들어가죠.
 - 15초 동안 147m를 갔을 때 속도는 <9.8m/s: QT_SPEED>입니다.

10.8.4. 주의 사항

- (태깅 대상 - 환산 수치) 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
 - 660.907127노트(1마하) → <660.907127노트: QT_SPEED>(<1마하: QT_SPEED>)
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 1마하(mach) → <1마하: QT_SPEED>(mach)

10.9. QT_TEMPERATURE

10.9.1. 정의 및 설명

- 온도, 기온, 수온 : 뜨거움과 차가움의 정도

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), 274.15절대온도(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 라뚠 온천이다.

10.9.4. 주의 사항

- (태깅 대상 - 환산 수치) 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>)
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 800도(°C) → <800도: QT_TEMPERATURE>(°C)

10.10. QT_VOLUME

10.10.1. 정의 및 설명

- 부피, 용량, 용적 : 물체나 물질이 공간에서 차지하는 크기

10.10.2. 분석 대상

- 부피, 용량, 용적
 - 2만 2000㎥, 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 따라 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.
 - <1fl.oz: QT_VOLUME>는 <약 29.57ml: 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

- (태깅 대상 - 환산 수치) 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
 - 1말(100홉) → <1말: QT_VOLUME>(<100홉: QT_VOLUME>)
- (태깅 대상 - 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 항공케널은 가로 350mm, 세로 500mm, 높이 650mm입니다. → 펫모드의 항공케널은 가로 <350mm: QT_LENGTH>, 세로 <500mm: QT_LENGTH>, 높이 <650mm: QT_LENGTH >입니다.
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 1갤런(gal) → <1갤런: QT_VOLUME>(gal)

10.11. QT_ORDER

10.11.1. 정의 및 설명

- 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

- n반: 3반, 4반 등
- n일차/n년차: 1일차, 8년차
- n주년/주기(周忌): 일 주년, 100주년, 백주년, 일 주기, 100주기, 백주기 등
- n기: 1기, 삼십오 기 등
- 목차: 1장, 2절, 3항 등
- n호/권/인쇄 출판: 제82호, 80호, 제8권, 10권, vol 2, 1쇄, 2쇄, 1권, 2판 등
- n회/회차: 제10회, 10회차 등
- D+n: D+234, D+6 등
- 원자 번호: 1번, 118번 등
- 특허 분야 도면과 청구항 번호, 특정한 기능 및 장치 번호: 도1, 도2, 청구항1 등
- 기타: 제4교구, 제2대, 12대, 제1대대, 일층, 3층, 5공구, 제7광구, 삼수, 4수, 4옥타브, 1코스, 2코스, 1구역, 2구역, 형사1부, 민사합의42부, 1호, 二號, 시즌1, 시즌2 등
- 등급
 - n위/등: 1위, 2등 등
 - n레벨/등급/단계: 90,000레벨, LV1, 1등급, 이 등급, 1단계, 십 단계 등
 - 한우 등급: 1, 2, 3등급, 1+, 1++, 1++A, 1++B, 1++C, 2A, 2B, 2C, 3A, 3B, 3C, 투뿔(투뿔) 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!
 - 서천초등학교 <6학년: QT_ORDER> <1반: QT_ORDER> 친구들의 사랑이 넘치는 가족공예체험수업!
 - 교생실습 <1일차: QT_ORDER>
 - 입사 <8년차: QT_ORDER>
 - 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 <백주년: QT_ORDER> 기념식이 서대문형무소에서 개최되었다.
 - 내일이 할아버지의 <이십오 주기: QT_ORDER>가 되는 날이다.
 - [하루 한 권 그림책] <17기: QT_ORDER> 모집 / 이번 기수 함께 읽을 엄마 책도 소개합니다.
 - 이 책의 체제 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(軍門謄錄), 난후잡록(亂後雜錄) 등으로 되어 있다.
 - 한국어 문법론의 개념어 연구 초판 <1쇄: QT_ORDER> 발행 2012년 11월 23일
 - 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 <형사1부: QT_ORDER>(부장 정준영)는 '회복적 사법' 개념을 적용한 판결로 사회에 파장을 일으켜 왔다.
- 초등학생 1학년 <D+6: QT_ORDER>, 오늘은 남산 나들이를 했습니다.
- 원자 번호 <1번: QT_ORDER> 수소(Hydrogen, H)부터 <118번: QT_ORDER> 오가네손(Oganesson, Og)까지 있다.
- <도1: QT_ORDER>, <도2: QT_ORDER>, <청구항1: QT_ORDER>, <청구항2: QT_ORDER>
- ~장치(<100: QT_ORDER>), ~모듈(<200: QT_ORDER>)
- <시즌1: QT_ORDER>이 <시즌2: QT_ORDER>보다 더 재미있다.
- 국보 <제4호: QT_ORDER> 여주 고달사지 승탑
- 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

- (태깅 대상 - 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.
- (태깅 대상 - 인쇄출판) 1쇄, 2쇄, 1권, 2판 등은 QT_ORDER로 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
 - 그는 전에 냈던 책의 증보판을 10년 만에 냈다. → 그는 전에 냈던 책의 증보판을 10년 만에 냈다.
- (태깅 대상 - 게임레벨) 레벨과 관련하여 '만렙'은 주의해서 태깅한다. 숫자로 쓴 LV10,000, 10,000레벨, 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!
 - 이 문장에서 '만렙'은 최대치를 의미하므로 태깅하지 않는다.
- (태깅 대상 - 특허 분야) 특허 분야의 도면, 청구항, 특정한 기능이나 장치 등 뒤에 붙는 번호도 QT_ORDER로 태깅한다.
 - 도1, 도2, 청구항1, 청구항2 → <도1: QT_ORDER>, <도2: QT_ORDER>, <청구항1: QT_ORDER>, <청구항2: QT_ORDER>

- ~장치(100), ~모듈(200) → ~장치(<100: QT_ORDER>), ~모듈(<200: QT_ORDER>)
- (태깅 대상 - n일차/n년차) 日次가 '날의 차례'를 의미하므로 DT가 아닌 QT_ORDER로 태깅한다.
 - 교생실습 1일차 → 교생실습 <1일차: QT_ORDER>
 - 입사 8년차 → 입사 <8년차: QT_ORDER>
- (태깅 대상 제외 - 첫+명사) '첫'은 관형사이므로 단독으로는 태깅하지 않으며, '첫+보통명사'도 태깅하지 않는다. 사전 등재 여부와 관련 없이 일괄적으로 태깅하지 않는다. 단, '첫+날짜' 관련 보통명사는 DT로 태깅한다.
 - 첫 여행, 첫 출발, 첫차 → 첫 여행, 첫 출발, 첫차
 - 첫날, 첫 주말 → <첫날: DT_DAY>, <첫 주말: DT_DURATION>
- (태깅 단위 - 과목명+n) 과목명은 FD로, 'n'은 QT_ORDER로 태깅한다. '생1, 생2'와 같이 줄임말로 나와도 '생물' 과목임을 알 수 있는 맥락이면 1, 2는 QT_ORDER로 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.
- (태깅 단위 - 목차) 장, 절, 항 등에 대해서는 각각 태깅하는 것을 기본으로 하되, 콜론(:)과 같은 기호와 같이 쓰인 경우에는 묶어서 태깅한다.
 - 시편 제69편 5절 → 시편 <제69편: QT_ORDER> <5절: QT_ORDER>
 - 시 69:5을 소리내어 읽어 보시다. → 시 <69:5: QT_ORDER>을 소리내어 읽어 보시다.
- (태깅 단위 - 학년+반) 'n학년 n반' 구성의 경우 학년과 반을 따로 태깅한다. '6-1'과 같은 'n-n'의 구성은 묶어 태깅한다.
 - 서천초등학교 6학년 1반 친구들의 사랑이 넘치는 가족공예체험수업! → 서천초등학교 <6학년: QT_ORDER> <1반: QT_ORDER> 친구들의 사랑이 넘치는 가족공예체험수업!
 - 현대고등학교 3-1 정미소의 가방. → 현대고등학교 <3-1: QT_ORDER> 정미소의 가방.
- (태깅 단위 - 보물, 국보, 세계문화유산+n호) '보물 1호, 국보 1호, 국보 제4호' 등 '보물, 국보, 세계문화유산+n호'가 나올 경우 '보물, 국보, 세계문화유산'은 태깅하지 않고 'n호'만 QT_ORDER로 태깅한다.
 - 국보 제4호 여주 고달사지 승탑 → 국보 <제4호: QT_ORDER> <여주 고달사지 승탑: AF_CULTURAL_ASSET>
- (태깅 단위 - 날짜+순서+상) 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>
- (태깅 범주 - n회) 회차를 의미하는 n회는 QT_ORDER로 태깅하지만, 횟수를 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"

- (태깅 범주 - ~째) '~째'가 '순서, 차례'의 의미일 때만 QT_ORDER, '동안'의 의미일 경우에는 DT나 T로 태깅한다.
 - 3일째 비가 오고 있었다. → <3일째: DT_DURATION> 비가 오고 있었다.
 - 3일째 되던 날, 비가 그쳤다. → <3일째: QT_ORDER> 되던 날, 비가 그쳤다.
- (태깅 범주 - n권) 발행 순서의 '권'은 QT_ORDER, 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.
- (태깅 범주 - D+n(일)) 'D+n(일)' 구성의 경우에는 QT_ORDER로 태깅하는 것을 원칙으로 하고, '생후 n 일'의 의미를 가질 때에 한해 QT_AGE로 태깅한다.
 - 초등학교 1학년 D+6, 오늘은 남산 나들이를 했습니다. → 초등학교 1학년 <D+6: QT_ORDER>, 오늘은 남산 나들이를 했습니다.
 - '생후'의 의미가 아닌 초등학교가 된 지 여섯 번째 날에 남산 나들이를 한 것이므로 순차로 보고 QT_AGE가 아닌 QT_ORDER로 태깅한다.
 - 빈나네 마당냥이 4남매 D+100!! 백일잔치~츄르잔치~ → 빈나네 마당냥이 4남매 <D+100: QT_AGE> 백일잔치~츄르잔치~
 - 문맥상 생후 n일째의 의미이므로 QT_AGE로 태깅한다.
- (태깅 대상 - 시즌n) 프로그램명 없이 '시즌n'이 나올 경우 QT_ORDER로 태깅한다. 단 '프로그램명+시즌n'이 함께 나올 경우, 전체를 AFA_VIDEO로 태깅한다.
 - 시즌1이 시즌2보다 낫다. → <시즌1: QT_ORDER>이 <시즌2: QT_ORDER>보다 낫다.
 - 모던패밀리 시즌1이 제일 좋았다. → <모던패밀리 시즌1: AFA_VIDEO>이 제일 좋았다.
- (태깅 범주 - N학기) 'N학기'가 학기의 순서, 순차를 나타내면 QT_ORDER, 학기를 세는 단위라면 QT_COUNT, '내내, 동안' 등과 함께 쓰여 지속성을 의미하는 것이라면 DT_DURATION으로 태깅한다.
 - 2학기에 세부 전공을 정해야 한다. → <2학기: QT_ORDER>에 세부 전공을 정해야 한다.
 - 일반적인 두 번째 학기를 지칭하기 때문에 QT_ORDER로 태깅한다.
 - 지금까지 총 4학기를 들었다. → 지금까지 <총 4학기: QT_COUNT>를 들었다.
 - 지금까지 총 4번을 들은 것으로 해석된다. 세는 단위이기 때문에 QT_COUNT로 태깅한다.
 - 2학기 내내 수학을 열심히 공부했다. → <2학기: DT_DURATION> 내내 수학을 열심히 공부했다.

10.12. QT_PRICE

10.12.1. 정의 및 설명

- 값, 가격, 금액 : 어떤 대상에 일정하게 매겨진 액수

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 26억 원: QT_PRICE> 되는 돈에 낙찰됐습니다.
 - 윤상현 씨가 전소속사로부터 <10억 원대: QT_PRICE>의 손해배상소송에 휩싸였습니다.

10.12.4. 주의 사항

- (태깅 대상 - 숫자만 나온 경우) 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

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
 - 1승 2패, n승/패/무, 5타점, 3:2, 2연승
 - n경기, 한 판, 두 판
 - n호골, n안타, 1볼넷, 6이닝, 6오버파

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.
 - 7월 21일 경기에서 SK 김광현이 <6이닝: QT_SPORTS> <2실점: QT_SPORTS>으로 승리 투수가 되었습니다.
 - <한 판: QT_SPORTS>만 더 이기면 강호동이 우승이다.

10.14.4. 주의 사항

- (태깅 대상 제외 - 스포츠 규칙, 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 박

석민이 <삼진: TM_SPORTS>으로 물러나면서 이닝이 종료되었다.

- 1레인이나 8레인처럼 가장자리에 있는 레인은 물의 저항을 많이 받아 불리할 수밖에 없다. → <1레인: TM_SPORTS>이나 <8레인: TM_SPORTS>처럼 가장자리에 있는 레인은 물의 저항을 많이 받아 불리할 수밖에 없다.
- (태깅 대상 제외 - n관왕) '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 바라보게 됐다. → 수원시청은 김나리-홍승연이 복식에도 결승에 올라, 단체전 우승에 이어 <3관왕: QT_COUNT>을 바라보게 됐다.
- (태깅 범주 - 한 판/한판) '수량+판'의 구성인 '한 판'은 QT_SPORTS로 태깅하고 유도의 판정 용어로 쓰인 '한판'은 TM_SPORTS로 태깅한다. 띄어쓰기를 참조하여 판단한다.
 - 한 판만 더 이기면 강호동이 우승이다. → <한 판만: QT_SPORTS> 더 이기면 강호동이 우승이다.
 - '두 판, 세 판'으로 교체가 가능하므로 이 문장에서 '한 판'은 '수량+판'의 구성이다. QT_SPORTS로 태깅한다.
 - 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.
 - 맥락상 유도의 판정 용어로 쓰였으므로 TM_SPORTS로 태깅한다.
- (태깅 범주 - QT_SPORTS / EV_SPORTS) '숫자+라운드, 회, 세트, 경기, 강'은 스포츠 행사의 경우 수량을 의미하면 QT_SPORTS, 경기 자체를 의미하면 EV_SPORTS로 태깅한다. '4강, 8강' 등과 같이 숫자가 나와도 문맥에 따라 QT가 아닌 EV로 처리해야 하는 경우가 있으므로 주의해야 한다.
 - 이번 K리그 축구는 총 32라운드를 경기할 예정이다. → 이번 K리그 축구는 <총 32라운드: QT_SPORTS>를 경기할 예정이다.
 - 라운드의 수를 나타내므로 QT_SPORTS로 태깅한다.
 - 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.
 - '3라운드'라는 경기 자체를 의미하므로 EV_SPORTS로 태깅한다.
 - 한화는 7회까지 <2회: QT_SPORTS>를 제외하고 매회 누상에 주자를 남겨둔 채 공격을 접어야 했다.
 - '7회 중 2회'의 의미로, 수량을 나타내므로 QT_SPORTS로 태깅한다.
 - 추신수는 9회 마지막 타석에서 헛스윙 삼진으로 물러났으나 텍사스는 1점 차로 이겼다. → 추신수는 <9회: EV_SPORTS> 마지막 타석에서 헛스윙 삼진으로 물러났으나 텍사스는 1점 차로 이겼다.
 - '9회 경기'를 나타내므로 EV_SPORTS로 태깅한다.
 - 아스날은 이번 경기에서의 승리로 칼링컵 포함 <14경기: QT_SPORTS>에서 13승 1무를 하며 무패 행진을 이어 가게 되었다.
 - '총 14경기'라는 수량을 나타내므로 QT_SPORTS로 태깅한다.
 - 프로야구는 6강전, 5강전을 진행한다. → 프로야구는 <6강전: EV_SPORTS>, <5강전: EV_SPORTS>을 진행한다.
 - '6강전, 5강전'은 경기 자체를 말하므로 EV_SPORTS로 태깅한다.

10.15. QT_CHANNEL

10.15.1. 정의 및 설명

- 텔레비전 · 라디오 · 무선 통신 따위의 주파수, 채널 번호

10.15.2. 분석 대상

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

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

- (태깅 대상 - 명확한 채널번호) 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 태깅한다.
 - 39쇼핑은 CJ오쇼핑에 인수되었습니다. → <39쇼핑: OGG_ECONOMY>은 <CJ오쇼핑: OGG_ECONOMY>에 인수되었습니다.
 - 문맥에서 채널명이 아닌 방송기관으로 판단될 경우 기관명에 해당하는 OGG_ECONOMY로 태깅한다.
- (태깅 대상 제외 - 채널명) KBS1TV, KBS2TV, EBS1, EBS2의 경우는 숫자가 포함되어 있어도 채널명에 해당되기 때문에 AFA_VIDEO로 태깅한다.
 - KBS1TV, KBS2TV, EBS1, EBS2 → <KBS1TV: AFA_VIDEO>, <KBS2TV: AFA_VIDEO>, <EBS1: AFA_VIDEO>, <EBS2: AFA_VIDEO>

10.16. QT_ALBUM

10.16.1. 정의 및 설명

- 노래 또는 연주곡 따위의 음반과 관련된 순차 표현

10.16.2. 분석 대상

- 테이프, 레코드판, 시디, 음원파일 등
 - 1집, 1.5집, VI, part1, 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

- (태깅 단위 - 앨범명+n집) 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

- 주소 관련 숫자 표현

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

- (태깅 단위 - 도로명 n길/n로 + 숫자) '도로명+n길, n로'는 QT_ADDRESS로 태깅하지 않는다. 도로명에 붙은 'n길', 'n로' 등은 도로명과 묶어서 AF_ROAD로 태깅하고 도로명 다음에 나오는 상세 번지와 건물 번호 및 동-호수는 QT_ADDRESS로 태깅한다.
 - 서울숲9길 9-7 → <서울숲9길: AF_ROAD> 9-7
 - 테헤란로 7길 8 → <테헤란로 7길: AF_ROAD> 8
- (태깅 단위 - 아파트명 n단지/n차/n동 + 숫자) 'n단지, n차, n동'은 단독으로 QT_ADDRESS로 태깅하지 않는다. 아파트명 등에 붙은 'n단지', 'n차'는 '아파트명+n단지/n차'+n동'과 함께 묶어 AF_BUILDING으로 태깅한다. 그 뒤에 나오는 호수는 QT_ADDRESS로 태깅한다. 단, 아파트명 등에 'n단지', 'n차'만 붙어 쓰였을 경우에는 LC_OTHERS로 태깅한다.
 - 잠실주공아파트 5단지 524동 802호 → <잠실주공아파트 5단지 524동: AF_BUILDING> 802호
 - 장미 2차 아파트 29동 505호 → <장미 2차 아파트 29동: AF_BUILDING> 505호
 - 레미안서초 6차에 전세 나온 거 있어요? → <레미안서초 6차: LC_OTHERS>에 전세 나온 거 있어요?
- (태깅 단위 - n-n) '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>
- (태깅 단위 - 주소 태깅)
 - (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>
 - 서울시 강남구 삼성로 1길 150 대치미도 1차 아파트 103-101 → 서울시 강남구 삼성로 1길 150 <대치미도 1차 아파트: AF_BUILDING> <103-101: QT_ADDRESS>
 - 공동주택의 동/호수를 'n-n'의 형식으로 묶어 쓴 경우에는 아래와 같이 QT_ADDRESS로 태깅한다.

10.18. QT_OTHERS

10.18.1. 정의 및 설명

- QT 계열의 세부 유형에 속하지 않는 기타 수량 표현

10.18.2. 분석 대상

- 기타 수량
 - 점수: 45점, 100점, 87/100(100점 만점에 87점의 의미)
 - 백분위: 백분위 등급 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
 - 수학적 기호가 포함된 경우: 1+1, 20×50sets, 51×70×47
 - +, -, ÷, ×, / 의 수학적 기호가 쓰인 경우, 맥락에 상관없이 QT_OTHERS로 태깅한다.
 - 비율이 아닌 n:n
 - 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 세 자리수가 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.
 - 모카프라프치노 <1+1: QT_OTHERS>행사 중

- 펫모드 스카이 항공켄넬 <51×70×47: QT_OTHERS>
- 한나라당이 <1:0: QT_OTHERS>로 이겼다.
- <1:1: QT_OTHERS> 상담을 신청했다.
- 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 <30,000,000 골드: QT_OTHERS>와 <11 루아의 증표: QT_OTHERS>를 보너스로 드립니다.

10.18.4. 주의 사항

- (태깅 대상 - 환산 수치) 괄호 안에 다른 단위로 환산된 수치가 표기되어 있을 경우에는 별도로 태깅한다.
 - 1hPa(100Pa:QT_OTHERS) → <1hPa: QT_OTHERS>(<100Pa: QT_OTHERS>)
- (태깅 대상 - 수학적 기호가 포함된 경우) +, -, ÷, ×, / 의 수학적 기호가 쓰인 경우, 수학적 기호를 포함한 수식이 분석 대상 문장만으로 특정 수치를 가리키면 QT_OTHERS로 분석하고, 그렇지 않으면 수식으로 간주하여 QT 태깅 대상에서 배제한다. 분석 문장에서 '숫자+수학적 기호+숫자'로 나올 경우, 결괏값이 확정되므로 전체를 QT_OTHERS로 태깅하고, '+3, x2'와 같은 '수학적 기호+숫자'의 경우, 결괏값이 정해지지 않으므로 태깅하지 않는다.
 - 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>
 - 2+2 총 4시간 → <2+2: QT_OTHERS> 총 4시간
 - 여기서 +2를 하는 겁니다.
 - '+2'는 '수학적 기호+숫자'기 때문에 QT로 태깅하지 않는다.
- (태깅 대상 제외 - 괄호 안 단위) 동일 단위에 대한 표기일 경우에는 괄호 안의 단위는 태깅하지 않는다.
 - 1기가바이트(GB) → <1기가바이트: QT_OTHERS>(GB)
- (태깅 대상 제외 - 정체를 알 수 없는 불명확한 숫자) 정체를 알 수 없는 불명확한 숫자는 태깅하지 않는다.
 - +254, #1#, ***12 등
- (태깅 단위 - 버스 노선) 노선버스의 번호에 붙는 영문은 태깅 범위에 포함한다.
 - 동탄2신도시 M버스 M4434, G버스 G6009 신안인스빌리베라 1차 정차! → 동탄2신도시 M버스 <M4434: QT_OTHERS>, G버스 <G6009: QT_OTHERS> 신안인스빌리베라 1차 정차!
- (태깅 범주 - n:n) 'n:n' 구성이 승부를 나타내거나, 사람의 수를 의미하는 경우에만 QT_OTHERS로 태깅한다. 비율을 나타낼 경우에는 QT_PERCENTAGE로 태깅한다.
 - 한나라당이 <1:0: QT_OTHERS>으로 이겼다.
 - <1:1: QT_OTHERS> 상담을 신청했다.
 - 사진을 <3:4: QT_PERCENTAGE> 비율로 조절했다.

11. EVENT

- EVENT는 특정 시간이나 장소에서 일어난 사건/사고/행사 명칭을 대상으로 한다.

분류 명칭	정의 및 설명
11.1. EV_ACTIVITY	- 사회운동 및 선언 명칭
11.2. EV_WAR_REVOLUTION	- 전쟁, 전투/작전, 반란/혁명/정변 명칭
11.3. EV_SPORTS	- 스포츠/레저 관련 행사 명칭 - e 스포츠 행사 명칭
11.4. EV_FESTIVAL	- 축제, 영화제, 콘서트, 박람회/엑스포, 행사 명칭 - 축제 및 행사의 보통명사
11.5. EV_OTHERS	- EV 계열의 세부 유형에 속하지 않는 기타 행사 : 사건/사고, 캠페인, 회의, 학술대회, 경진대회, 전시회 등

- 주의 사항
 - (태깅 대상 - 보통명사) EV_SPORTS와 EV_FESTIVAL은 보통명사도 태깅한다. 나머지 EV는 시기, 기간이 명확한 구체적인 사건에 태깅한다.

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, 갑오경장, 항일독립운동
- 선언
 - 인권선언, 노예해방선언

11.1.3. 예문

- 사회운동
 - 1962년 흑인의 <투표권등록운동: EV_ACTIVITY>이 시작되었다.
 - 1931년부터 1934년까지 동아일보사가 전개한 <브나로드운동: EV_ACTIVITY>은 고등보통학교 4, 5학

년 학생으로 이루어진 학생계몽대와 전문학교 이상의 학생으로 조직된 학생강연대, 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.
- 조선왕조는 <갑오경장: EV_ACTIVITY>이라는 이름의 개혁을 통해 쓰러져가는 국운을 되살리려 애썼다.
- 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

- (태깅 대상 제외 - OO 칙령) '칙령'의 사전적 의미는 '임금이 내린 명령'으로 선언을 의미하기도 하지만, 발표, 명령, 선포 등 다양한 의미를 가지므로 EV_ACTIVITY로 태깅하지 않고 CV_LAW로 태깅한다.
 - 밀라노 칙령에는 그리스도교 공인에 관한 획기적인 명령이 들어있습니다. → <밀라노 칙령: CV_LAW>에는 그리스도교 공인에 관한 획기적인 명령이 들어있습니다.
- (태깅 대상 제외 - 사보타주) 사보타주는 노동운동보다는 특정 행동으로 볼 수 있으므로 태깅하지 않는다.
 - UPS의 사보타주를 보상하기 위해 무선 이어폰을 선물했다.

11.2. EV_WAR_REVOLUTION

11.2.1. 정의 및 설명

- 전쟁
- 전투/작전
- 반란/혁명/정변

11.2.2. 분석 대상

- 전쟁
 - 러일전쟁, 남북전쟁, 포클랜드전쟁, 아편전쟁, 임진왜란, 명량해전, 칠천량해전, 노량해전, 연평해전, 2차세계대전
- 전투/작전
 - 봉오동 전투, 황산벌 전투, 위털루 전투, 노르망디상륙작전, 인천상륙작전
- 반란/혁명/정변
 - 청교도혁명, 4.19 혁명, 신해혁명, 메이지유신, 갑신정변, 516 쿠데타, 임술민란, 임술농민봉기, 정여립 모반 사건, 이인좌의 난, 인조반정

11.2.3. 예문

- 전쟁
 - <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.
 - <2차세계대전: EV_WAR_REVOLUTION>은 20세기 전 인류에 최고의 영향을 미친 사건으로 기록될 것입니다.
- 전투/작전
 - 당연히 <안시성 전투: EV_WAR_REVOLUTION>는 고구려-당나라 전쟁에서 최대의 클라이맥스가 됩니다.
 - 김홍춘의 아들인 반굴은 김유신과 부친과 함께 <황산벌 전투: EV_WAR_REVOLUTION>에 참전했다.
 - <노르망디상륙작전: EV_WAR_REVOLUTION>의 성공으로 연합군은 그해 8월 25일 파리를 탈환하였다.
- 반란/혁명/정변
 - <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.
 - 한국 현대사의 비극중의 하나인 <516 쿠데타: EV_WAR_REVOLUTION>가 1961년 5월 16일 새벽에 발생했다.
 - 박규수는 1862년 <임술민란: EV_WAR_REVOLUTION> 때는 백성을 무마하는 안핵사로 진주에 내려갔었다.
 - 1884년 12월 <갑신정변: EV_WAR_REVOLUTION> 때, 민영익과 그 동료들이 큰 부상을 입게 된다.
 - <임술농민봉기: EV_WAR_REVOLUTION>는 단성에서 시작되고 곧 이어 진주로 파급되어갔다.
 - <정여립 모반 사건: EV_WAR_REVOLUTION>으로 인한 기축옥사로 동인이 대거 제거된다.
 - 청주에서 <이인좌의 난: EV_WAR_REVOLUTION>이 일어난 것은 영조 4년(1728)이었다.
 - 1623년 <인조반정: EV_WAR_REVOLUTION> 성공 나흘 뒤 왕위에 오른 능양군 이종이 선언했다.

11.2.4. 주의 사항

- (태깅 대상 제외 - 프롤레타리아혁명) 시기를 특정할 수 없는 개념에 해당하므로 태깅하지 않는다.
 - 프롤레타리아혁명 실험은 실패로 결론났다.
- (태깅 대상 제외 - 사회변화) 시기, 기간을 특정할 수 없는 사조나 사회변화를 의미하는 OO혁명은 CV_CULTURE로 태깅한다.
 - n차산업혁명, 르네상스, 농업혁명

11.3. EV_SPORTS

11.3.1. 정의 및 설명

- 스포츠 관련 행사
- e 스포츠 행사

11.3.2. 분석 대상

- 스포츠 관련 행사
 - 마라톤 대회, 등반 대회, 축구 대회, 자전거 대회, 테니스 대회, 태권도 대회, 조정 대회, 홈경기, 원정 경기
 - 서울 올림픽, 88 서울올림픽, 거제시장기 전국 윈드서핑대회, 세계 역도 선수권 대회, 남아공월드컵 축구대회, 내셔널리그 챔피언십시리즈
 - 두산전, 개막전, 최종전, 결승전, 예선전, 전반기, 후반기, 조별예선전, 주중 3연전, A매치, 호주전, 한국-호주전
- 게임 관련 e 스포츠 이벤트(행사) 이름
 - 오버워치 리그, 던전앤파이터 프리미어 리그, LoL 월드챔피언십

11.3.3. 예문

- 스포츠 관련 행사
 - 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!
 - 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.
 - <전반기: EV_SPORTS>, <후반기: EV_SPORTS> 풀타임을 소화한 손흥민 선수
- e 스포츠 행사
 - 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 수닝을 꺾고 우승했다.

11.3.4. 주의 사항

- (태깅 대상 - n파전) EV_SPORTS로 태깅한다.
 - 고양시 女축구대회, 한중일 3파전으로 열린다. → 고양시 女축구대회, 한중일 <3파전: EV_SPORTS>으로 열린다.
- (태깅 대상 - 줄임말) 문맥을 고려하여 태깅한다.
 - 평창 티켓 있으신가요? 20일 쇼트트랙 → <평창: EV_SPORTS> 티켓 있으신가요? 20일 쇼트트랙
 - 88 서울 때 마스코트는 호돌이다. → <88 서울: EV_SPORTS> 때 마스코트는 호돌이다.
- (태깅 단위 - 제n회+OO 올림픽) 개최 회차는 분리하여 태깅하고 지역은 통합하여 태깅한다.
 - 제24회 서울 올림픽 경기 → <제24회: QT_ORDER> <서울 올림픽 경기: EV_SPORTS>
- (태깅 단위 - 연도+스포츠 행사) 스포츠 행사가 연도와 병행하여 쓰일 경우 명확한 연도 표시는 DT_YEAR로 태깅하고, 축약된 경우는 통합하여 EV_SPORTS로 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 타고르고 있다. → <88 서울올림픽경기: EV_SPORTS> 16일간의 대장정을 알리는 성화가 막 점화되어 타고르고 있다.
- (태깅 단위 - n회 말, n회 초) '3회 말', '9회 초' 등의 경우 '말, 초'를 제외하고 '3회, 9회'만 따로 EV_SPORTS로 태깅한다.
 - 진짜 승부는 9회 말부터였다. → 진짜 승부는 <9회: EV_SPORTS> 말부터였다.
- (태깅 범주 - EV_SPORTS/QT_SPORTS) '숫자+라운드, 회, 세트, 경기, 강'은 스포츠 행사의 경우 '4강, 8강' 등과 같이 숫자가 나와도 문맥에 따라 QT가 아닌 EV로 처리해야 하는 경우가 있으므로 주의해야 한다.
 - <KPGA 오픈 3라운드: EV_SPORTS>에서 김민규가 1위에 오르며 지난주 군산CC 오픈에서 우승한 김주형에 이어 10대 돌풍을 이어갈 기세다.
 - 이번 K리그 축구는 <총 32라운드: QT_SPORTS>를 경기할 예정이다.
 - '32라운드'는 32번의 경기를 의미하므로 QT_SPORTS로 태깅한다.
 - 추신수는 <9회: EV_SPORTS> 마지막 타석에서 헛스윙 삼진으로 물러났으나 텍사스는 1점 차로 이겼다.
 - 한화는 7회까지 <2회: QT_SPORTS>를 제외하고 매회 누상에 주자를 남겨둔 채 공격을 접어야 했다.
 - '2회'는 1~7회까지의 공격 중 두 번의 헛수를 말하므로 QT_SPORTS로 태깅한다.
 - 역전 드라마는 마지막 <5세트: EV_SPORTS>에서 펼쳐졌다.
 - T1은 결승전에서 <2세트: QT_SPORTS>를 먼저 선취하며 우승까지 단 1승을 남겨두었다.
 - '2세트'는 세트의 개수를 의미하므로 QT_SPORTS로 태깅한다.
 - 이 사진은 2019 <세계 남자농구월드컵 조별예선 3경기: EV_SPORTS>의 하이라이트 장면이다.
 - 아스날은 이번 경기에서의 승리로 칼링컵 포함 <14경기: QT_SPORTS>에서 13승 1무를 하며 무패 행진을 이어 가게 되었다.
 - '14경기'는 '아스날'이 치른 경기의 수를 말하므로 QT_SPORTS로 태깅한다.
 - 프로야구는 <6강전: EV_SPORTS>, <5강전: EV_SPORTS>을 진행한다.
 - 이번 시즌에는 KT까지 <6강: QT_SPORTS>에 포함된다.
 - '6강'은 6개의 강팀을 의미하므로 QT_SPORTS로 태깅한다.

- (태깅 범주 - 유로) '유로'는 스포츠 행사라면 EV_SPORTS, 화폐 명칭이라면 CV_CURRENCY로 태깅한다.
 - 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.

11.4. EV_FESTIVAL

11.4.1. 정의 및 설명

- 축제 명칭
- 영화제 명칭
- 콘서트 명칭
- 박람회/엑스포 명칭
- 행사 명칭
- 축제 및 행사의 보통명사

11.4.2. 분석 대상

- 축제 명칭
 - 갯방구대축제, 새미골막사발축제, 이천도자기축제, 수원화성문화제, 보령머드축제, 담양대나무축제, 부산바다축제, 부천국제만화축제, 울산서머페스티벌
 - ~대(학교) 축제, ~대동제
- 영화제 명칭
 - 베니스 국제영화제, 칸 영화제, 부산국제영화제, 부천영화제
- 콘서트 명칭
 - BOYSWISH 콘서트, 방탄 콘서트, 트와이스 콘서트, 나훈아 콘서트
- 박람회/엑스포 명칭
 - 스마트국토엑스포, GTI 국제무역투자박람회, 서울커피엑스포, 제일 창업박람회
- 행사 명칭
 - 책 읽는 춘천, 2019 서울모터쇼, 코리아 세일 페스타, 서울 패션 위크, 코베 베이비페어
- 축제 및 행사의 보통명사
 - 연어축제, 장미축제, 연등축제, 단풍축제, 록 페스티벌

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.
- 박람회/엑스포 명칭
 - 온라인으로 열린 <스마트국토엑스포: EV_FESTIVAL> 행사에서 전시 관계자들이 온라인 전시 영상을 제작하고 있다.
 - 밀키트창업, 밀키트가게가 대세였던 2021 <제일창업 박람회: EV_FESTIVAL>
- 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.
- 축제 및 행사의 보통명사
 - 연어를 맨손으로 안아보는 체험은 <연어축제: EV_FESTIVAL>의 백미다.
 - 삼광사 <연등축제: EV_FESTIVAL>가 유명하다고 한다.

11.5. EV_OTHERS

11.5.1. 정의 및 설명

- EV 계열의 세부 유형에 속하지 않는 기타 행사
 - 사건/사고
 - 역사적인 사건/사고

- 캠페인
- 사태
- 회의
- 학술대회(회의)
- 전시회, 영화 시사회, 팬사인회
- 시험/경진대회
- 축제 내 체험행사

11.5.2. 분석 대상

- EV 계열의 세부 유형에 속하지 않는 기타 행사
 - 사건/사고: 세월호 참사, 태안 기름유출사고, 국정농단, 이메일 게이트, 플라자 합의, 포츠머스조약, 남북 평화협정
 - 역사적인 사건/사고: 위화도 회군, 헤이그특사 사건, 헤이그밀사 사건, 카노사의 굴욕, 삼전도의 굴욕
 - 캠페인: 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
 - 전시회, 영화 시사회, 팬사인회: 키스해링전, 체르노빌 상영회, 베컴 팬사인회, ART TOY CULTURE 2014 SEOUL, 샤갈 러브 앤 라이프전, 기생충 상영회, 어벤져스 상영회, 트와이스 팬사인회
 - 시험/경진대회: 국제뷰티기능대회, 청소년저작권글짓기대회, 해군호국문예제, 하나은행 자연사랑 어린이 미술대회, 미스유니버스, 월드바리스타챔피언쉽
 - 축제 내 체험행사: 연어맨손잡기체험, 대나무체험행사, 카누체험, 북 쥐띠 문양 페이스페인팅

11.5.3. 예문

- EV 계열의 세부 유형에 속하지 않는 기타 행사
 - <세월호 참사: EV_OTHERS>는 2014년 4월 16일 인천에서 제주로 향하던 중에 발생하였다.
 - 온 국민을 안타깝게 했던 <태안기름유출사고: 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.
 - 그런데 미국은 이번 <페르시아만 사태: EV_OTHERS>를 자신들의 뜻대로 해결했을 경우, 소련을 국제 무대에서 확실하게 따돌릴 수 있다고 확신하는 듯하다.
 - 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'.
- 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.
- 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 17일까지 만나실수 있습니다.
- 전편 무료 <체르노빌 상영회: EV_OTHERS>를 10월 20일 롯데시네마 월드타워 수퍼플렉스G에서 개최한다고 합니다.
- 29일 오후 <베컴 팬사인회: EV_OTHERS>는 몰려든 인파로 인한 안전상의 문제로 시작 15분 만에 취소됐다.
- 2019 <국제뷰티 기능대회: EV_OTHERS> 출전 생생 후기 포스팅하려 합니다!
- <월드바리스타챔피언십: EV_OTHERS>이 오는 2017년 11월 9일부터 12일까지 개최된다.
- 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 윷을 던져 새해 운수를 점치는 <윷점 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

- (태깅 대상 - 선거, 주민투표) 문맥상에서 일시, 장소가 명확한 특정 행사, 사건으로 볼 수 있는 항목인 경우에만 태깅한다. '-시행하다, -진행하다' 등 적정 술어와 공기한다.
- 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 서울시장의 무상급식에 대한 <주민투표: EV_OTHERS>를 통해 투표율 33.3%를 달성하지 못함에 따라 사퇴하였다.
- (태깅 대상 - OO정상회담, OO협상) 문맥상에서 일시, 장소가 명확한 구체적인 사건으로 특정할 수 있

는 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.
- (태깅 대상 - 수능, 임용고시) 특정 시간에 특정 장소에서 이루어지는 사건일 경우에만 태깅한다. '-시행되다, -실시하다, -치르다' 등의 적정 술어와 공기하는 경우 EV_OTHERS로 태깅한다.
 - 배우들도 멋있고 수능 공부로 쌓인 스트레스가 확 날아갈 만큼 역동적이고 재미있는 공연이었다.
 - 2021 임용고시는 일정대로 11월 21일(토)에 시행했다. → 2021 <임용고시: EV_OTHERS>는 일정대로 11월 21일(토)에 시행했다.
- (태깅 단위 - 날짜+횟수+행사명) 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'.
- (태깅 범주 - OO조약, OO협정) 문맥에 준하여 EV_OTHERS와 CV_LAW로 태깅한다. '-을 체결하다, -을 맺다' 등의 적정 술어와 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

- ANIMAL은 진핵생물 중에서 동물에 해당되는 것으로, '사람'을 제외한 짐승을 대상으로 한다.

분류 명칭	정의 및 설명
12.1. AM_INSECT	- 곤충 명칭
12.2. AM_BIRD	- 조류 명칭 - 조류의 하위 유형
12.3. AM_FISH	- 어류 명칭 - 어류의 하위 유형
12.4. AM_MAMMALIA	- 포유류 명칭
12.5. AM_AMPHIBIA	- 양서류 명칭
12.6. AM_REPTILIA	- 파충류 명칭
12.7. AM_TYPE	- 동물 분류 명칭
12.8. AM_PART	- 동물이나 사람 몸의 한 부분(신체 부위) 명칭
12.9. AM_OTHERS	- AM 계열의 세부 유형에 속하지 않는 기타 동물 명칭

- 주의 사항
 - (태깅 대상 - 천연기념물) 천연기념물로 지정된 동물명은 그에 해당하는 AM 계열로 태깅한다.
 - 추애산조선범의 표지주는 고산군 죽근리 봉수령에 있다. → <추애산조선범: AM_MAMMALIA>의 표지주는 고산군 죽근리 봉수령에 있다.
 - (태깅 범주 - 음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 AM 계열이나 CV_FOOD로 태깅한다.
 - 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.
 - 고등어는 따뜻한 바다에서 서식한다. → <고등어: AM_FISH>는 따뜻한 바다에서 서식한다.
 - 오늘 저녁에는 고등어로 요리를 해야겠다. → 오늘 저녁에는 <고등어: CV_FOOD>로 요리를 해야겠다.

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.

12.1.4. 주의 사항

- (태깅 범주 - 절지동물류) 곤충은 머리-가슴-배의 구분이 있어 절지동물에 속하나 따로 분류해 AM_INSECT로 태깅한다. 곤충이 아닌 절지동물은 AM_OTHERS로 태깅한다.
 - 개미는 턱이 발달되어 튼튼하다. → <개미: AM_INSECT>는 턱이 발달되어 튼튼하다.
 - 특히 사막에서는 전갈을 조심해야 한다. → 특히 사막에서는 <전갈: AM_OTHERS>을 조심해야 한다.

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, 뺨꾸기, 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

- (태깅 대상 제외 - 새 이름+알) '새 이름+알'의 경우 복합어이므로 AM_BIRD의 태깅하지 않는다.
 - 달걀, 계란, 메추리알
- (태깅 범주 - 음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 AM_BIRD나 CV_FOOD로 태깅한다.
 - 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.

12.3. AM_FISH

12.3.1. 정의 및 설명

- 어류의 이름
- 어류의 하위 유형

12.3.2. 분석 대상

- 어류의 이름
 - 참가자미, 장어, 갯장어, 붕장어, 전기뱀장어, 가물치, 고등어, 방어, 연어, 갈치, 멸치, 미꾸리, 송어, 날치, 붕어, 버들붕어, 콩치, 돔, 돌돔, 참돔, 감성돔, 아귀, 초롱아귀 등
- 어류의 하위 유형
 - 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

- (태깅 대상 제외 - 어류 이름+알) '어류 이름+알'의 경우 복합어이므로 AM_FISH로 태깅하지 않는다.
 - 청어알, 상어알, 날치알
- (태깅 범주 - 음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 AM_FISH나 CV_FOOD로 태깅한다.
 - 고등어는 따뜻한 바다에서 서식한다. → <고등어: AM_FISH>는 따뜻한 바다에서 서식한다.
 - 오늘 저녁에는 고등어로 요리를 해야겠다. → 오늘 저녁에는 <고등어: CV_FOOD>로 요리를 해야겠다.

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

- (태깅 대상 - 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.
- (태깅 대상 제외 - 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.

- (태깅 대상 제외 - 인간) 사람, 인간 및 인류 진화의 개체는 태깅하지 않는다.
 - '인간, 사람, 호모사피엔스, 오스트랄로피테쿠스' 등은 태깅하지 않는다.

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)·<꼬리치레도롱뇽: AM_AMPHIBIA>(Onychodactylusfischeri)·<네발가락도롱뇽: AM_AMPHIBIA>(Salamandrella keyserlingi)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
 - 우리나라에는 <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
 - 200여 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

- (태깅 대상 - 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

- 신체 부위나 동물 몸 부위의 명칭
 - 신체 부위
 - 동물 몸 부위

12.8.2. 분석 대상

- 신체 부위나 동물 몸 부위의 명칭

- 신체 부위: 팔, 다리, 머리, 장딴지, 허벅지, 팔꿈치, 눈, 코, 입, 이, 혀, 머리카락, 눈썹, 피부
- 동물 몸 부위: 팔, 다리, 머리, 대가리, 지느러미, 부리, 날개, 꼬리, 이빨, 혀, 털, 피부

12.8.3. 예문

- 신체 부위나 동물 몸 부위의 명칭
 - 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.
 - 갖가지 화려한 색상을 뿜어내는 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

- (태깅 대상 - 이, 털) 신체 조직이나 기관 중에서 눈으로 보이는 부분만 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
- (태깅 대상 - 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.
- (태깅 대상 제외 - 뼈) 눈으로 확인할 수 있는 부위만 AM_PART로 태깅한다.
 - 골반이 틀어졌다. → <골반: TM_CELL_TISSUE_ORGAN>이 틀어졌다.
 - '골반'은 뼈를 지칭하는 것이기 때문에 AM_PART로 태깅하지 않는다.
- (태깅 대상 제외 - 껍질/껍데기) '껍질', '껍데기'는 물질을 나타내므로, AM_PART에 해당하지 않는다.
 - 고등어 껍질을 제거해야 비린 맛이 덜 난다.
- (태깅 대상 제외 - 신체 부위+국소 지칭) 특정 신체 부위의 일부분을 지칭하는 '입꼬리, 손끝, 발끝, 눈가, 입가' 등과 같은 경우는 태깅하지 않는다.
 - 쇼파 밑에 있는 돈을 손끝으로 겨우 꺼냈다.
- (태깅 단위 - 방향+신체 부위) '방향+신체 부위'는 따로 태깅하지 않고 통합하여 태깅한다.
 - 왼발을 오른 다리 위에 놓고 → <왼발: AM_PART>을 <오른 다리: AM_PART> 위에 놓고
- (태깅 단위 - 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.

12.9. AM_OTHERS

12.9.1. 정의 및 설명

- AM 계열의 세부 유형에 속하지 않는 기타 동물의 이름
 - 절지동물류
 - 연체동물류
 - 기타 생물류
 - 미생물류
 - 상상 속 동물류

12.9.2. 분석 대상

- AM 계열의 세부 유형에 속하지 않는 기타 동물의 이름
 - 절지동물류: 가재, 새우, 대하, 토하, 게, 대게, 꽃게, 지네
 - 연체동물류: 오징어, 문어, 낙지, 달팽이, 민달팽이, 조개, 굴, 전복
 - 기타 생물류: 해파리, 산호, 멍게, 해삼, 불가사리, 말미잘, 지렁이
 - 미생물류: 세균, 바이러스, 박테리아, 아메바
 - 상상 속 동물류: 유니콘, 용, 해태, 봉황, 삼족오

12.9.3. 예문

- AM 계열의 세부 유형에 속하지 않는 기타 동물의 이름
 - <가재: AM_OTHERS>는 <게: AM_OTHERS>편이라고, 감사합니다.
 - <붉은대게: AM_OTHERS>는 1,000m 바다 밑 청정지역에 삽니다.
 - <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.
 - 슬리피 라군이라 불리는 <산호초: AM_OTHERS>들이 파도를 막아 주면서 자연적으로 형성된 이스트 아가나 비치에 있는 호텔이다.
 - 뛰어난 항균력으로 <곰팡이: AM_OTHERS>나 <박테리아: AM_OTHERS>는 물론 오염에도 강한 내구성이 좋은 소재예요!
 - 드래곤보트 위에서는 그들 모두가 다 광저우의 <용: AM_OTHERS> 아닐까요?

12.9.4. 주의 사항

- (태깅 범주 - 음식) 개체가 동물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 AM_OTHERS나 CV_FOOD로 태깅한다.
 - 민물 새우를 토하라고 합니다. → <민물 새우: AM_OTHERS>를 <토하: AM_OTHERS>라고 합니다.
 - 대하, 맛있죠. → <대하: CV_FOOD>, 맛있죠.
- (태깅 범주 - 미생물류) 미생물류는 문맥에 따라 태깅한다.
 - 김치 유산균을 배양해 만든 소금으로 → 김치 <유산균: AM_OTHERS>을 배양해 만든 소금으로

- 유산균을 주기적으로 복용하면 장 건강에 좋다. → <유산균: TMM_DRUG>을 주기적으로 복용하면 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.

분류 명칭	정의 및 설명
13.1. PT_FRUIT	- 과일, 식물의 열매/씨앗
13.2. PT_FLOWER	- 꽃 이름
13.3. PT_TREE	- 나무 이름
13.4. PT_GRASS	- 풀이름
13.5. PT_TYPE	- 식물 분류 명칭
13.6. PT_PART	- 식물의 조직, 부분의 명칭
13.7. PT_OTHERS	- PT 계열의 세부 유형에 속하지 않는 기타 식물 명칭

- 주의 사항
 - (태깅 대상 - 천연기념물) 천연기념물로 지정된 식물명은 그에 해당하는 PT 계열로 태깅한다.
 - 원산금송은 금슬가운데서 제일 큰것중의 하나이다. → <원산금송: PT_TREE>은 금슬가운데서 제일 큰것중의 하나이다.
 - (태깅 대상 제외 - OO내, OO취, OO겹질, OO발) <우리말샘>에 '-' 또는 '^' 기호와 함께 등재되어 있는 경우, 복합어에 준하는 것으로 보아 복합어 내부의 구성요소는 PT 계열로 태깅하지 않는다.
 - 자몽내, 사과내, 사과취, 굴겹질, 사과겹질 등
 - 딸기발, 감자발 등
 - (태깅 범주 - 음식) 개체가 식물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 PT 계열이나 CV_FOOD로 태깅한다.
 - 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.
 - 양배추는 발육과정이 복잡하다. → <양배추: PT_GRASS>는 발육과정이 복잡하다.
 - 양배추를 조금 먹어선 크게 효과를 볼 수 없어요. → <양배추: CV_FOOD>를 조금 먹어선 크게 효과를 볼 수 없어요.
 - (태깅 범주 - 다의어 식물) 식물명은 식물의 일부분을 가리키기도 하고 식물 전체를 가리키기도 하는 등 다의성을 가지는데, 적정 술어를 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.

13.1. PT_FRUIT

13.1.1. 정의 및 설명

- 과일, 식물의 열매/씨앗

13.1.2. 분석 대상

- 과일
 - 토종 과일: 감, 구기자, 머루, 모과, 무화과, 산사자, 산수유, 오미자 등
 - 열대 과일: 바나나, 파인애플, 아보카도, 키위, 망고, 망고스틴, 코코넛, 두리안, 롱안, 리치, 패션프루트 등
 - 굴류 과일: 귤, 감귤, 레몬, 라임, 오렌지, 자몽, 유자 등
 - 딸기류 과일: 블루베리, 체리, 크랜베리, 라즈베리, 복분자딸기, 서양복분자, 딸기, 곰딸기, 명석딸기 등
 - 핵과류 과일: 대추, 매실, 복숭아, 살구, 앵두, 자두 등
- 식물의 열매/씨앗
 - 견과류 열매: 땅콩, 밤, 아몬드, 잣, 피스타치오, 헤이즐넛, 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

- 과일
 - 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 들어온 것으로 추측되며 《조선왕조실록》 등에도 <포도: PT_FRUIT>에 관한 기록이 실려 있다.
- 식물의 열매/씨앗
 - <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

- (태깅 단위 - 과일명+향/추출물) <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
- 치자 추출물 → <치자: PT_FRUIT> 추출물
- 이 음료에는 석류 추출물이 들어 있다. → 이 음료에는 <석류: PT_FRUIT> 추출물이 들어 있다.
- (태깅 범주 - 음식) 개체가 식물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 PT 계열이나 CV_FOOD로 태깅한다.
 - 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.

13.2. PT_FLOWER

13.2.1. 정의 및 설명

- 꽃의 이름

13.2.2. 분석 대상

- 꽃의 이름
 - 봄꽃: 개나리, 진달래, 제비꽃, 짙레꽃, 벚꽃, 튤립, 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: 갯쭉부쟁이, 국화, 겹삼입국화, 감국, 산국, 코스모스, 억새꽃, 메밀꽃, 백일홍, 채송화 등
 - 겨울꽃: 동백, 수선화, 군자란, 프리지어, 칼랑코예, 복수초 등
 - 그 외 꽃: 데이지, 안개꽃, 백합, 히아신스, 개꽃 등

13.2.3. 예문

- 꽃의 이름
 - <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

- (태깅 단위 - 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>
- (태깅 범주 - 다의어 식물) 식물명은 식물의 일부분을 가리키기도 하고 식물 전체를 가리키기도 하는 등 다의성을 가지는데, 적정 술어를 이용해서 하위의미를 구분하여 태깅한다.
 - 산수유가 노랑게 피었다. → <산수유: PT_FLOWER>가 노랑게 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

- 나무의 이름

13.3.2. 분석 대상

- 나무의 이름
 - 침엽수: 삼나무, 소나무, 잣나무, 전나무, 분비나무, 가문비나무, 종비나무, 잎갈나무, 주목, 편백나무, 낙엽송, 낙우송, 눈잣나무, 눈측백나무, 솔송나무, 향나무, 구상나무, 풍산가문비나무 등
 - 활엽수: 참나무, 벚나무, 호도나무, 물푸레나무, 피칸, 오리나무, 느티나무, 단풍나무, 목련나무, 피나무, 오동나무, 떡갈나무, 뽕나무, 상수리나무, 동백나무, 너도밤나무 등

13.3.3. 예문

- 나무의 이름
 - 한국에서는 <후박나무: PT_TREE> 숲을 생활환경으로 하고 있어서 <후박나무: PT_TREE>와 흑비둘기의 분포가 밀접한 관계를 가지고 있으며
 - 부인 일가는 묵은 <벚나무: PT_TREE>들이 짙혀 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

- (태깅 대상 - 천연기념물) '양평 용문사 은행나무'와 같은 천연기념물도 PT_TREE로 태깅한다.
 - 양평 용문사 은행나무 → 양평 용문사 <은행나무: PT_TREE>
- (태깅 대상 - 단풍) '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.
 - 온 산에 빨강고 노랑게 단풍이 들었다.

- 여기서 '단풍'은 나무를 가리키지 않으므로 태깅하지 않는다.
- (태깅 대상 제외 - 재료) 다음과 같이 나무가 재료로서 쓰인 경우에는 태깅하지 않는다.
 - 편백나무로 책상을 만들었다.

13.4. PT_GRASS

13.4.1. 정의 및 설명

- 풀의 이름

13.4.2. 분석 대상

- 풀의 이름
 - 엽채류: 갓, 근대, 머위, 무청, 미나리, 배추, 상추, 셀러리, 시금치, 싹갓, 양배추 등
 - 경채류: 두릅, 마늘종, 부추, 셀러리, 아스파라거스, 엄교, 죽순, 파 등
 - 근채류: 당근, 래디시, 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 풀이름: 토끼풀, 향부자, 하늘타리, 갈대, 질경이, 강아지풀, 창포, 억새, 잔디, 담배, 파슬리, 한란 등

13.4.3. 예문

- 채소의 이름
 - 이거 완전 <인삼: PT_GRASS> 뿌리같이 생겼는데요.
 - <고구마: PT_GRASS> 수확이 오늘 미션입니다.
 - <미나리: PT_GRASS>는 봄이 제철이다.
- 풀의 이름
 - 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.
 - 화분에 심은 <파슬리: PT_GRASS>는 매일 햇빛을 쏘도록 해 줘야 잘 자란다.
 - <한란: PT_GRASS>은 한국에서는 제주도의 한라산에서 자라며, 천연기념물로서 보호받고 있다.

13.4.4. 주의 사항

- (태깅 대상 - 담배) 담배 (연초)는 문맥에서 명확하게 식물, 담뱃잎을 의미하는 경우에 해당되었을 때만 PT_GRASS로 태깅한다. 그 외의 경우에는 태깅하지 않는다.
 - 올해는 장마로 담배농사를 망쳤다. → 올해는 장마로 <담배: PT_GRASS> 농사를 망쳤다.

- 벌레가 담배의 잎을 갉아먹었다. → 벌레가 <담배: PT_GRASS>의 잎을 갉아먹었다.
- (태깅 범주 - 음식) 개체가 식물의 의미로 쓰였을 때와 음식으로 쓰였을 때를 구분하여 PT 계열이나 CV_FOOD로 태깅한다.
 - 양배추는 발육과정이 복잡하다. → <양배추: PT_GRASS>는 발육과정이 복잡하다.
 - 양배추를 조금 먹어선 크게 효과를 볼 수 없어요. → <양배추: CV_FOOD>를 조금 먹어선 크게 효과를 볼 수 없어요.

13.5. PT_TYPE

13.5.1. 정의 및 설명

- 식물 분류의 명칭

13.5.2. 분석 대상

- 식물 분류의 명칭
 - 녹조류, 선태식물, 양치식물, 양치류, 종자식물
 - 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
 - 콩과, 참나무과, 생강과, 박과

13.5.3. 예문

- 식물 분류의 명칭
 - <박과: PT_TYPE>의 덩굴성 식물인 동아는 손부가 직접 기르고 수확한 것입니다.
 - <녹조류: PT_TYPE>는 엽록소를 가지고 있다.
 - <콩과: PT_TYPE>의 열매는 대부분 떡잎이 크게 발달한 것이다.

13.5.4. 주의 사항

- (태깅 대상 제외 - 보통명사) 아래 예와 같은 식물 보통명사 명칭은 PT_TYPE으로 태깅하지 않는다.
 - 저지대 식물, 고산식물, 꽃, 나무, 열대 과일, 야생화 등
- (태깅 대상 제외 - OO류) 'OO류'로 식물 보통명사 명칭을 나타낼 경우 PT_TYPE으로 태깅하지 않는다.
 - 과실류, 견과류, 곡류, 버섯류 등
- (태깅 단위 - 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

- 식물의 조직, 부분의 명칭

13.6.2. 분석 대상

- 식물의 조직, 부분의 명칭
 - 싹, 나무갓수관, 뿌리, 뿌리목, 얇은 뿌리, 결뿌리, 뿌리털
 - 가지, 큰가지, 결가지, 잔가지, 가시, 열매, 잎, 줄기, 밑줄기
 - 꽃, 꽃잎, 꽃봉오리, 꽃수술, 꽃받침, 씨앗 등

13.6.3. 예문

- 식물의 조직, 부분에 대한 명칭
 - <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 좋고 하는데요.

13.6.4. 주의 사항

- (태깅 대상 - 꽃) 보통명사 '꽃'과 식물의 부분으로서의 '꽃' 모두 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.
- (태깅 대상 제외 - 껍질/껍데기) '껍질'은 식물에만 존재하는 것이 아닌 일반 물질이므로 PT_PART로 태깅하지 않는다.
 - 후박나무 껍질, 버드나무 껍질 등
- (씨앗 이름) '해바라기씨', '목화씨'는 PT_PART로 태깅한다.
 - 해바라기씨 → <해바라기씨: PT_PART>
 - 목화씨 → <목화씨: PT_PART>

13.7. PT_OTHERS

13.7.1. 정의 및 설명

- PT 계열의 세부 유형에 속하지 않는 기타 식물의 이름
 - 이끼류
 - 버섯류

- 해조류
- 기타 식물류
- 기타 식물의 하위 유형 명칭

13.7.2. 분석 대상

- PT 계열의 세부 유형에 속하지 않는 기타 식물의 이름
 - 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
 - 기타 식물류: 선인장, 콩나물 등
- 기타 식물의 하위 유형 명칭
 - 이끼, 버섯, 해초 등

13.7.3. 예문

- PT 계열의 세부 유형에 속하지 않는 기타 식물의 이름
 - <상황버섯: PT_OTHERS>에 상어지느러미까지.
 - 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.
- 기타 식물의 하위 유형 명칭
 - 사실 갯지렁이는 모랫속에 들어 있는 미생물과 <해초: PT_OTHERS>를 먹고 삽니다.
 - <버섯: PT_OTHERS>은 독이 있기도 하다.

14. MATERIAL

- MATERIAL은 원소, 금속, 암석, 화학물질 등의 물질명을 대상으로 한다.

분류 명칭	정의 및 설명
14.1. MT_ELEMENT	- 원소명 - 준금속 원소명
14.2. MT_METAL	- 금속물 명칭
14.3. MT_ROCK	- 암석류/보석류/광석류 명칭
14.4. MT_CHEMICAL	- 화학 물질 명칭

- 주의 사항
 - (태깅 단위 - 복합어 내부 태깅) MT 계열은 해당 원소명 및 금속물 등 물질의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다. <우리말샘> 사전에 '^'로 연결되어 있는 경우, MT에 한해 '^'를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)한다.
 - 수소강국 위한 어벤져스 탄생 → <수소: MT_ELEMENT>강국 위한 어벤져스 탄생
 - 마그네슘리본 → <마그네슘: MT_METAL>리본
 - 칼슘왕 미꾸라지 → <칼슘: MT_METAL>왕 미꾸라지
 - 산소마스크 → 산소마스크
 - 금비녀 → 금비녀
 - '산소마스크', '금비녀'는 <우리말샘>에 '산소-마스크', '금-비녀'로 등재되어 있으므로 단어 속 MT 계열의 물질 의미가 유지되는 것으로 보여도 태깅하지 않는다.

14.1. MT_ELEMENT

14.1.1. 정의 및 설명

- 원소명
 - 원소기호와 번호를 할당받은 주기율표에 등록된 물질
- 준금속 원소명

14.1.2. 분석 대상

- 원소명
 - 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨
- 준금속 원소명
 - 붕소, 규소, 게르마늄(저마늄), 비소, 안티모니, 텔루륨, 폴로늄

14.1.3. 예문

- 원소명
 - 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 산업용전기로 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 데크항공의 대주주인 (주)데크와 체결했다.
- 준금속 원소명
 - <게르마늄: MT_ELEMENT>으로 키운 농산물
 - 순수한 <비소: MT_ELEMENT>는 독성이 없다.

14.1.4. 주의 사항

- (태깅 대상 - 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>
 - 이때의 NA는 나트륨의 원소기호로 쓰인 것이 아닌 North America의 약자이므로 MT로 태깅하지 않는다.
- (태깅 단위 - 복합어 내부 태깅) MT_ELEMENT는 해당 원소명 및 금속물 등 물질의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다. <우리말샘> 사전에 '^'로 연결되어 있는 경우, MT에 한해 '^'를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)한다.
 - 수소강국 위한 어벤져스 탄생 → <수소: MT_ELEMENT>강국 위한 어벤져스 탄생
 - 산소마스크
 - '산소마스크'는 <우리말샘>에 '산소-마스크'로 등재되어 있으므로 단어 속 MT_ELEMENT의 물질 의미가 유지되는 것으로 보여도 태깅하지 않는다.
- (태깅 범주 - 금속 원소) 금속 물질 중에서 금속 원소는 MT_ELEMENT와 MT_METAL이 모두 태깅 가능한데, 이때는 MT_METAL로 태깅한다.
 - 리튬, 알루미늄, 구리, 철, 아연, 은, 금, 납 등
 - 티타늄 재질이라 가볍다. → <티타늄: MT_METAL> 재질이라 가볍다.
- (태깅 범주 - 원소 화학물질) 화학물질 중에 원소에 해당하는 경우에는 MT_ELEMENT와 MT_CHEMICAL

이 모두 가능한데, 원소명에 해당하는 경우에는 무조건 MT_ELEMENT로 태깅한다.

- 산소 생산 해양 박테리아 → <산소: MT_ELEMENT> 생산 해양 박테리아

14.2. MT_METAL

14.2.1. 정의 및 설명

- 금속물 명칭

14.2.2. 분석 대상

- 금속물 명칭
 - 금, 은, 동, 칼슘, 황금, 황산철, 알루미늄, 아연, 구리, 수은, 아말감

14.2.3. 예문

- 금속물 명칭
 - 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

- (태깅 대상 제외 - 준금속 원소) 준금속 원소는 MT_ELEMENT로 태깅한다.
 - 붕소, 규소, 게르마늄(저마늄), 비소, 안티모니, 텔루륨, 폴로늄
 - 게르마늄으로 키운 농산물 → <게르마늄: MT_ELEMENT>으로 키운 농산물
- (태깅 단위 - 복합어 내부 태깅) MT_METAL은 해당 원소명 및 금속물 등 물질의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다. <우리말샘> 사전에 '^'로 연결되어 있는 경우, MT에 한해 '^'를 기준으로 띄어쓰기 관계없이 내부 태깅(분리 태깅)한다.
 - 마그네슘리본 → <마그네슘: MT_METAL>리본
 - 칼슘왕 미꾸라지 → <칼슘: MT_METAL>왕 미꾸라지
 - 금비녀 → 금비녀
 - '금비녀'는 <우리말샘>에 '금-비녀'로 등재되어 있으므로 단어 속 MT_METAL의 물질 의미가 유지되는 것으로 보여도 태깅하지 않는다.
- (태깅 범주 - 금속 원소) 금속 물질 중에서 금속 원소는 MT_METAL과 MT_ELEMENT이 모두 태깅 가능한데, 이때는 MT_METAL로 우선 태깅한다.

- 리튬, 알루미늄, 구리, 철, 아연, 은, 금, 납 등
- 티타늄 재질이라 가볍다. → <티타늄: MT_METAL> 재질이라 가볍다.

14.3. MT_ROCK

14.3.1. 정의 및 설명

- 암석류의 명칭
- 보석류의 명칭
- 광석류의 명칭

14.3.2. 분석 대상

- 암석류의 명칭
 - 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암, 석탄, 연탄, 무연탄, 황토
- 보석류의 명칭
 - 사파이어, 다이아몬드, 루비, 진주, 크리스탈, 옥
- 광석류의 명칭
 - 철광석, 금광석, 은광석, 동광석

14.3.3. 예문

- 암석류의 명칭
 - 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.
 - 산책로에는 목재 데크와 <황토: MT_ROCK> 포장을 사용해 쉽게 다닐 수 있도록 하고 주변 경관과 어울릴 수 있도록 할 계획입니다.
- 보석류의 명칭
 - 파란색 <사파이어: MT_ROCK>와 <다이아몬드: MT_ROCK>로 된 약혼반지
 - <옥: MT_ROCK>과 같은 고가의 재료도 사용된 만큼 기린상은 중국인들에게 인기가 높다.
 - 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.
- 광석류의 명칭
 - 실제로 인도산 <철광석: MT_ROCK>의 중국향 현물가격(스팟가격)은 지난 6월 t당 124달러에서 8월 149달러로 20%이상 상승했고 최근 오퍼가격 역시 155달러까지 올랐다.
 - 호주서 56억 원어치 거대 <금광석: MT_ROCK> 발견

14.3.4. 주의 사항

- (태깅 단위 - 암석 명칭+~류) 암석 명칭의 경우 '~류' 등과 결합하여 암석 종류를 나타낼 때 암석만을 태깅 대상으로 한다.
 - 화강암류 → <MT_ROCK: 화강암>류

14.4. MT_CHEMICAL

14.4.1. 정의 및 설명

- 화학물질의 명칭

14.4.2. 분석 대상

- 화학물질의 명칭
 - 고체 화학물질: 파라핀, 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, 단백질 등

14.4.3. 예문

- 화학물질의 명칭
 - <탄산나트륨: MT_CHEMICAL>을 이용하여 제조한 순환잔골재의 품질 특성
 - <나프탈렌: MT_CHEMICAL>을 대체할 천연 탈취제를 찾고 계신가요?
 - <수산화나트륨: MT_CHEMICAL>과 <탄산칼륨: MT_CHEMICAL> 수용액을 전기분해한 알칼리 이온수
 - 한국석유공사 등에 따르면 국제 유가 하락 등에 힘입어 21일 0시를 기해 제주지역 기름값이 <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
 - 화장실에서 <암모니아: MT_CHEMICAL> 냄새가 나는 원인은?
 - 배달대행업체 '전기이륜차' 도입 <이산화탄소: MT_CHEMICAL> 배출 줄인다.
 - <CO2: MT_CHEMICAL>를 줄이는 탄소포인트제예요.
 - 보리수 열매에는 <아미노산: MT_CHEMICAL>과 <비타민C: MT_CHEMICAL>가 풍부해서 기관지에는 물론 피부미용에도 좋다고 하네요.
 - 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

- (태깅 대상 - 식물 화학물질) 식물에서 나오거나 식물이 유발하는 화학물질은 MT_CHEMICAL로 태깅한다.
 - 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.
- (태깅 대상 제외 - 식물 화학물질) 식물의 어떤 부분이나 그러한 명칭 중에 물질로 보기 어려운 것은 MT_CHEMICAL로 태깅하지 않는다.
 - 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.
- (태깅 대상 제외 - OO 추출물, OO 향) "석류 추출물"과 같은 추출물 부류는 그 추출물 전체를 물질의 이름으로 보기 어려우므로 전체를 MT_CHEMICAL로 태깅하지 않는다. "사과향", "오렌지향"과 같은 향의 부류도 이와 비슷한 경우이다.
 - 치자 추출물 → <치자: PT_FRUIT> 추출물
 - 망고 추출물 → <망고: PT_FRUIT> 추출물
 - 사과 향 → <사과: PT_FRUIT> 향
 - 오렌지 향 → <오렌지: PT_FRUIT> 향
 - 복합어가 아닌 경우에는 과일명은 PT_FRUIT로 태깅한다.
- (태깅 범주 - 원소 화학물질) 화학물질 중에 원소에 해당하는 경우에는 MT_CHEMICAL과 MT_ELEMENT 모두 가능한데, 원소명에 해당하는 경우에는 무조건 MT_ELEMENT로 태깅한다.
 - 산소생산 해양 박테리아 → <산소: MT_ELEMENT>생산 해양 박테리아
- (태깅 범주 - 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

15. TERM

- TERM은 위에서 정의된 세부 개체명 이외의 개체명 용어를 대상으로 한다.

분류 명칭	정의 및 설명
15.1. TM_COLOR	- 색 및 빛의 이름
15.2. TM_DIRECTION	- 방위, 방향 - 기타 방향 표현
15.3. TM_CLIMATE	- 기후 분류 명칭
15.4. TM_SHAPE	- 모양의 유형 : 형태, 무늬
15.5. TM_CELL_TISSUE_ORGAN	- 세포 명칭 - (인체 내의) 장기, 기관 및 조직 명칭(외부에서 보이지 않는 조직이나 기관)
15.6. TMM_DISEASE	- 증상 및 질환
15.7. TMM_DRUG	- 약 및 의약품명 : 외용약, 내복약, 기타 의약품
15.8. TMI_HW	- IT 하드웨어 명칭 및 용어 - 전자제품 - 장치명, 장치를 구성하는 부품 - 로봇
15.9. TMI_SW	- IT 소프트웨어 용어 - 기타 소프트웨어 용어
15.10. TMI_SITE	- URL 주소
15.11. TMI_EMAIL	- 이메일 주소
15.12. TMI_MODEL	- 각종 제품의 세부 모델명, 모델일련번호 - 예) LM1000-2CX
15.13. TMI_SERVICE	- IT 서비스 용어
15.14. TMI_PROJECT	- 프로젝트 명칭
15.15. TMIG_GENRE	- 게임 장르 명칭
15.16. TM_SPORTS	- 스포츠/레저 용어(기술/규칙 명칭)

15.1. TM_COLOR

15.1.1. 정의 및 설명

- 색 및 빛의 이름

- 계통색 이름
- 관용색 이름
- 빛 이름
- 톤 이름

15.1.2. 분석 대상

- 색 및 빛의 이름
 - 계통색 이름: 갈색, 갈회색, 갈흑색, 검은색, 남색, 노란색, 녹색, 녹회색, 보라색, 분홍색, 빨간색, 연노란색, 연두색, 연분홍색, 자주색, 적자색, 주황색, 진분홍색, 청록색, 초록색, 파란색, 황갈색, 회갈색, 회색, 회적색, 회청색, 흰색 등
 - 관용색 이름: 계란색, 국방색, 당근색, 대추색, 딸기색, 루비색, 물색, 박하색, 벚꽃색, 벽돌색, 복숭아색, 사과색, 산호색, 살구색, 석류색, 수박색, 썩색, 연지색, 우유색, 장미색, 코코아색, 토마토색, 풀색, 하늘색, 호두색, 호박색, 홍차색 등
 - 빛 이름: 갈색빛, 구릿빛, 개나리빛, 검은빛, 노란빛, 녹색, 다홍빛, 보랏빛, 은빛, 핏빛, 흰빛 등
 - 톤 이름: 화이트 톤, 블루 톤, 핑크 톤, 레드 톤, 그레이 톤, 그린 톤 등

15.1.3. 예문

- 색 및 빛의 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.
 - <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.
 - 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.
 - 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 대상 - 줄임말) '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!

- (태깅 대상 - 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.
- (태깅 대상 제외 - 임의로 붙인 색상명) 화장품 제조사에서 특정 색상에 임의로 붙인 이름은 TM_COLOR로 태깅하지 않는다.
 - 입생로랑 워터티트 '바그드 루쥬'와 '인베리딕'은 말을 더해 줘도 부족한 예쁜 컬러입니다. → 입생로랑 워터티트 '바그드 루쥬'와 '인베리딕'은 말을 더해줘도 부족한 예쁜 컬러입니다.
 - '바그드 루쥬, 인베리딕, 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'은 블랙핑크 지수와 제니가 화보 속에서 착용한 렌즈이다.
 - 화장품 외에도 렌즈, 휴대전화 등 제조사에서 임의로 붙인 색상명은 TM_COLOR로 태깅하지 않는다.
- (태깅 단위 - 형용사 결합형) 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.
- (특수 항목 태깅 - 기호 삽입) 기호가 삽입되어 있는 경우에는 기호를 포함하여 태깅한다.
 - 올 가을은 팔죽+포도주스색이 인기를 끌고 있습니다. → 올 가을은 <팔죽+포도주스색: TM_COLOR> 이 인기를 끌고 있습니다.

15.2. TM_DIRECTION

15.2.1. 정의 및 설명

- 방향
- 방위
- 기타 방향 표현

15.2.2. 분석 대상

- 방향
 - 상측, 하측, 상향, 서북향, 아래쪽, 앞, 앞뒤, 앞쪽, 옆쪽, 오른쪽, 오른쪽, 왼쪽, 왼쪽, 우측, 위쪽, 동남

향, 뒤쪽, 남향, 좌측, 하향

- 방위
 - 동, 서, 남, 북, 중앙, 북서, 북동, 남서, 남동, 동서, 동북
- 기타 방향 표현
 - 지역의 부분: 서부, 남부, 동부, 북부, 중부, 최남단, 최동단, 최북단, 최서단
 - 위치: 위, 아래, 옆
 - 시계 관련 방향: 시계 방향, 반시계 방향, 7시 방향, 12시 방향

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.
- 기타 방향 표현
 - 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!
 - 책상 <위: TM_DIRECTION>에는 책이 있고, 책상 <아래: TM_DIRECTION>에는 가방이 놓여 있다.
 - <시계 방향: TM_DIRECTION>으로 순서가 돌아갑니다.
 - 다운스윙을 <3시 방향: TM_DIRECTION>으로 내리는 스윙

15.2.4. 주의 사항

- (태깅 대상 - 방향 합성어) '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 제외 - 방향성 이동 동사) 명사나 명사구만을 태깅 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.
- (태깅 대상 제외 - O부 지역/O부권) '북부권'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.

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

- 기후 분류 명칭
 - 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 제외 - 기후 특성) 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 표현은 태깅하지 않는다.
- (태깅 대상 제외 - 지대) '열대, 아열대, 온대' 등이 문맥상 '지리, 지대'를 의미할 때는 태깅하지 않는다.
 - 열대는 <열대기후: TMI_CLIMATE> 지역으로서 고온다습하고 고무, 코코야자, 코코아 등의 플랜테이션이 영위되고 있다.

15.4. TM_SHAPE

15.4.1. 정의 및 설명

- 모양의 유형
 - 형태
 - 무늬

15.4.2. 분석 대상

- 모양의 유형
 - 형태: 유선형, 일자형, 원뿔꼴, 원뿔형, 일자, 원, 삼각형, 사각형, 직사각형, 팔각형, 오각형, 피라미드형, 방울 패턴, 기하학적 패턴, 타원, 동그라미, 둥그러미, 둥근형, 선형, 소용돌이, 수상형, 실꼴, 십자꼴, 십자형, 이삭꼴, 피라미드형, 달걀모양, 둥근모양, 뚜껑모양, 타원모양, 사다리꼴모양, H형, 일자형, 원통형, 삼각뿔 모양, 사각뿔 모양
 - 무늬: 가로무늬, 구름무늬, 줄무늬, 점무늬

15.4.3. 예문

- 모양의 유형
 - <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?
 - 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 대상 - '형/모양/패턴' 생략) '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.
- (태깅 단위 - OO형/OO모양/OO패턴/무늬) 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.

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

- 세포의 명칭
 - 극세포, 간상세포, 딸세포, 신경세포, 정원세포, 정모세포, 가시세포, 각막세포, 그물세포, 주세포, 황색 세포, 유전자, 아교세포
- (인체 내의) 장기, 기관 및 조직 명칭
 - 근육, 혈액, 침샘, 뉴런, 망치뼈, 모루뼈, 등자뼈, 갈비뼈, 늑연골, 반지연골, 방패연골, 췌기연골, 복사 뼈, 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 늘어서 있으나, <간상세포: TM_CELL_TISSUE_ORGAN>의 수가 훨씬 더 많다.
 - 분열 전의 <모세포: TM_CELL_TISSUE_ORGAN>(母細胞)에 대해서 <딸세포: TM_CELL_TISSUE_ORGAN>라고 하며 <낭세포: TM_CELL_TISSUE_ORGAN>(娘細胞)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 분류 된다.
 - <표피: TM_CELL_TISSUE_ORGAN>의 주요 기능은 식물체의 보호이다.
 - <정원세포: TM_CELL_TISSUE_ORGAN>는 유사분열을 거듭해서 일정한 수준까지 증식을 한 다음 성장 과 성숙을 거쳐서 <정모세포: TM_CELL_TISSUE_ORGAN>(精母細胞)가 된다.
- (인체 내의) 장기, 기관 및 조직 명칭
 - 운동 후 마시는 프로틴이 맛도 좋고 <근육: TM_CELL_TISSUE_ORGAN>에도 좋다.
 - 최근 오뚜기에서는 <침샘: TM_CELL_TISSUE_ORGAN>을 자극하는 다양한 신제품을 출시했다.
 - <고막: TM_CELL_TISSUE_ORGAN>에 도달한 음파는 <고막: TM_CELL_TISSUE_ORGAN>을 진동시키는 데 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 수용하고 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 대상 제외 - 눈에 보이는 신체 부위) 눈으로 보이지 않는 부분만 TM_CELL_TISSUE_ORGAN으로 태깅한다. 눈에 보이는 '피부, 살, 이, 혀' 등의 부분은 AM_PART로 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.
 - 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.
- (태깅 대상 제외 - 혈관) '동맥', '정맥', '모세혈관'은 태깅 대상이나, '혈관'은 태깅하지 않는다.
- 3혈관은 동맥, 모세혈관, 정맥을 총칭한다. → 혈관은 <동맥: TM_CELL_TISSUE_ORGAN>, <모세혈관: TM_CELL_TISSUE_ORGAN>, <정맥: TM_CELL_TISSUE_ORGAN>을 총칭한다.
- (태깅 대상 제외 - 신체 기관 총칭) 기관을 총칭하는 '비뇨기' 등은 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.
- (태깅 단위 - 복합어) '소 간, 손가락 뼈' 등 띄어쓰기와 상관없이 사전에 등재되어 있는 것은 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?

15.6. TMM_DISEASE

15.6.1. 정의 및 설명

- 증상
- 질환

15.6.2. 분석 대상

- 증상
 - 기침, 가래, 경련, 딸꾹질, 발열, 재채기, 변비, 저체온증, 편두통, 멍, 출혈

- 질환
 - 내과 질환: 역류성식도염, 직장암, 폐결핵, 각기병, 위장염, 하혈, 늑막염, 패혈증, 식중독
 - 외과 질환: 콜린성두드러기, 헤르페스, 기면증, 뇌경색, 뇌출혈, 뇌진탕, 치매, 찰과상, 골절
 - 산부인과 및 소아과 질환: 구루병, 소아마비
 - 안과 및 치과 질환: 안검염, 결막염, 콩다래끼, 녹내장
 - 이비인후과 질환: 이하선염, 인후염, 고막염, 인후암, 후두염
 - 정신 질환: 고소공포증, 우울증, 강박증, 조현병, 분노조절장애
 - 기타 질환: 페스트, 스페인독감, 후유증, 알코올 중독, 니코틴 중독, 스마트폰 중독, 게임 중독

15.6.3. 예문

- 증상
 - 나이가 들면 <기침: TMM_DISEASE>을 하는 것이 당연하다고 생각하면 안 된다.
 - 거짓말을 하면 <말국질: 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.
 - 현대인의 절반 이상이 <우울증: TMM_DISEASE>, <강박: TMM_DISEASE> 등 <정신 질환: TMM_DISEASE>을 앓고 있는 것으로 조사되었다.

15.6.4. 주의 사항

- (태깅 대상 - 바이러스) 질병의 원인을 의미하는 '바이러스'는 TMM_DISEASE로 태깅한다.
 - 어린 아이들은 바이러스에 감염되기 쉽기에 특별히 조심해야 한다. → 어린 아이들은 <바이러스: 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.
- (태깅 대상 - OO 질환) '체장 질환, 심장 질환' 등은 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.

- (태깅 대상 제외 - 세균) '세균'은 증상, 질환의 원인으로 볼 수 있으나 증상, 질환이 아니므로 AM_OTHERS로 태깅한다.
 - 뇌수막염은 세균이 원인이기에 항생제가 필수이다. → <뇌수막염: TMM_DISEASE>은 <세균: AM_OTHERS>이 원인이기에 항생제가 필수이다.
- (태깅 대상 제외 - OO병) 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
- (태깅 단위 - 질환의 전체 명칭) 질환의 전체 명칭을 나누어 태깅할 경우, 질환의 의미가 달라질 수 있으므로 최장 단위 원칙에 따라 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.

15.7. TMM_DRUG

15.7.1. 정의 및 설명

- 약 및 의약품명
 - 외용약
 - 내복약
 - 기타 의약품

15.7.2. 분석 대상

- 약 및 의약품명
 - 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 의약품: 모기향, 생맥산

15.7.3. 예문

- 약 및 의약품명
 - 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(鰵) 외에, 건조갑 상선·경랍(鯨蠟)·취장·건조분말·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)는 충분한 양을 제때에 복용한다.

15.7.4. 주의 사항

- (태깅 대상 - 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.
 - '~을 짓다', '~을 조제하다', '~을 처방하다' 등과 같이 적정 술어가 사용되었을 때만 TMM_DRUG로 태깅한다.
- (태깅 대상 제외 - 의약 상품명) 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 상처에는 <후시딘연고: AFW_OTHER_PRODUCTS>와 같이 쉽게 찾아볼 수 있는 <연고: TMM_DRUG>를 바르면 된다.
- (태깅 단위 - OO 예방 영양제) 'OO 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.
- (태깅 범주 - 영양제 부류) 화학물질 하위 부류 중 영양성분 및 호르몬은 영양제 MT_CHEMICAL로도 태깅 가능한데, 문맥과 적정 술어를 통해 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
- (태깅 범주 - 유산균) 미생물에 속하는 유산균은 '복용하다, 먹다'와 같은 술어와 함께 쓰이면 AM_OTHERS가 아닌 TMM_DRUG로 태깅한다.
 - 김치 유산균을 배양해 만든 소금으로 → 김치 <유산균: AM_OTHERS>을 배양해 만든 소금으로
 - 유산균을 주기적으로 복용하면 장 건강에 좋다. → <유산균: TMM_DRUG>을 주기적으로 복용하면 장 건강에 좋다.

15.8. TMI_HW

15.8.1. 정의 및 설명

- 하드웨어 명칭 및 용어
- 전자제품
- 장치명
- 장치를 구성하는 부품
- 로봇

15.8.2. 분석 대상

- 하드웨어 명칭 및 용어
 - 입력장치: 키보드, 마우스, 스캐너, 마이크, 조이스틱
 - 중앙처리장치: 마이크로프로세서, 그래픽 프로세서, 사운드 프로세서, 메인 보드, 그래픽 카드, 사운드 카드, 랜 카드
 - 기억장치: 임시기억장치(램, 시스템 메모리, 캐시 메모리), 영구기억장치(롬), 플로피 디스크
 - 저장장치: 하드디스크 드라이브
 - 출력장치: 모니터, 프린터, 스피커
 - 코드: 바코드, QR코드
- 전자제품
 - 컴퓨터, 모니터, 팩스, 카메라, 디카, 디지털카메라, 냉장고, 전자레인지, 핸드폰, 마우스, 전화(기), 노트북, 계산기, 라디오, 에어컨, 세탁기, 텔레비전, 히터기, 내비게이션, 방수 카메라, 충전기, 스마트폰, 사운드바, 디지털 오디오, 엘피, 수화기, 카세트, 게임기
- 장치명
 - 망원경, 인공위성, 이동통신단말기, 무선화상수신부, 회동지지바, 연동지지바, 광섬유격자, 광전송선, 광서클레이터, 광검출부, 충격측정부
- 장치를 구성하는 부품
 - 소자 소켓, 감각 센서, 후각 센서, 감지 센서, LED
- 로봇
 - 뷰리오시티, 오퍼튜니티(화성 탐사 로봇), 스피릿(화성 탐사 로봇)

15.8.3. 예문

- 하드웨어 명칭 및 용어
 - <하드디스크: TMI_HW>는 자성체를 입힌 원판형 알루미늄 기판을 회전시키면서 자료를 저장하고 읽어 내도록 한 보조기억장치의 하나로써
 - <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)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.
- 계산대의 여종업원은 너무나 지루하다는 듯이 물건들의 <바코드: TMI_HW>를 기계에 읽힌 다음 비닐봉지에 담아 주었다.
- 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.
- 로봇
 - <오퍼튜니티: TMI_HW>는 2004년부터 활동한 화성 탐사 <로버: TMI_HW>다.

15.8.4. 주의 사항

- (태깅 대상 - OO장치/OO모듈/OO부) 특히 문서에는 특성상 장치명, 모듈명 등이 존재하는데, 개체명 대상이 아니지만 특히 분야에서는 중요한 키워드가 되기에 'OO장치, OO모듈, OO부' 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.
- (태깅 대상 제외 - 상품명) 전자제품 중 정확한 상품명으로 나온 경우만 AFW_OTHER_PRODUCTS로 태깅하고 그 외 보통명사에 해당하는 제품들은 TMI_HW로 태깅한다.
 - 아이폰3GS는 한국의 스마트폰 시장을 흔들어 놓았다. → <아이폰3GS: AFW_OTHER_PRODUCTS>

는 한국의 <스마트폰: TMI_HW> 시장을 흔들어 놓았다.

- (태깅 대상 제외 - 모델명) '카메라'와 같은 제품 유형은 TMI_HW로, 세부 제품 모델명은 TMI_MODEL로 태깅한다.
 - 초소형 슬라이드 카메라폰 E800이 사람들의 주목을 받고 있다. → <초소형 슬라이드 카메라폰: TMI_HW> <E800: TMI_MODEL>이 사람들의 주목을 받고 있다.

15.9. TMI_SW

15.9.1. 정의 및 설명

- 소프트웨어 명칭
 - 시스템 소프트웨어
 - 언어 번역 프로그램
 - 유틸리티 프로그램
 - 응용 소프트웨어
- 기타 소프트웨어 용어
 - 악성 코드
 - 프로그래밍 언어

15.9.2. 분석 대상

- 소프트웨어 명칭
 - 시스템 소프트웨어: 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 뎀, 익스플로러, 크롬, 오토캐드, 일러스트레이터, 워드 뷰어
- 기타 소프트웨어 용어
 - 악성 코드: 트로이목마, 웜 바이러스, 님다바이러스
 - 프로그래밍 언어: 자바, 파이선, C언어, 코볼

15.9.3. 예문

- 소프트웨어 명칭
 - 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(170여 개 기관)에 침투를 시도했다"고 밝혔다.
 - 회사 컴퓨터 <알집: TMI_SW>을 업데이트하면서 악성 프로그램을 내려받게 되었다.
 - 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.
- 기타 소프트웨어 용어
 - <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.
 - <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

- (태깅 범주 - 시스템 및 프로그램) 특허 문서에는 특성상 '~시스템, ~프로그램' 등이 존재하는데, 개체명 대상이 아니나 중요한 키워드기에 TMI_SW로 태깅한다. 단, '~시스템'은 문맥상 소프트웨어 프로그램이 분명할 때만 TMI_SW로 태깅하고, 어떠한 운용 방법을 의미할 때는 TR_SCIENCE로 태깅한다.
 - 지난 1월부터 중소기업에 적합한 명함관리 프로그램을 개발했다 → 지난 1월부터 중소기업에 적합한 <명함관리 프로그램: TMI_SW>을 개발했다
 - '데몬, 명함 관리 시스템, 위성위치확인시스템, 지리정보시스템' 등도 TMI_SW로 태깅한다.
- (태깅 대상 제외 - 인터넷 관련 보통명사) '사이트, 홈페이지, 링크' 등과 같은 인터넷 관련 보통명사는 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.
- (태깅 범주 - TMI_SW/AFW_OTHER_PRODUCTS) AFW_OTHER_PRODUCTS와 중복되는 경우 TMI_SW로 우선적으로 태깅한다.
 - 윈도우, 한컴오피스 한글, MS 오피스 등

15.10. TMI_SITE

15.10.1. 정의 및 설명

- URL 주소
 - http
 - https
 - 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

- URL 주소
 - 반려동물을 맡겨둘 수 있는 영업장은 동물보호관리시스템(<www.animal.go.kr: TMI_SITE>)에서 확인하시기를 바랍니다.
 - 문화재청은 국가문화유산포털(<http://www.heritage.go.kr: TMI_SITE>)의 문화재 정보 접근성과 검색 방식 등을 개선하여 이달부터 시행한다.
 - 국립국어원(<https://www.korean.go.kr: TMI_SITE>) 온라인가나다를 통해 어문 규범이나 어법 등을 문의할 수 있다.
 - '<mailto:admin@test.it: TMI_SITE>'와 같이 국세청을 사칭한 메일주소를 주의하시기 바랍니다.
 - 마이크로소프트 사의 FTP 주소는 '<ftp://ftp.microsoft.com: TMI_SITE>'입니다.

15.10.4. 주의 사항

- (태깅 대상 - URL 일부 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

- 이메일 주소
 - 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

- 제품의 모델명
 - 예약판매하는 그램 17 대표 모델(모델명: <17Z90N-VA50K: TMI_MODEL>) 출하가는 214만 원이다.
 - 프로기아가 상급자용 드라이버 <TR500: TMI_MODEL>을 출시했다.

15.12.4. 주의 사항

- (태깅 대상 제외 - 제품 유형) TMI_HW와 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.
- (태깅 대상 제외 - 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
 - 애플리케이션
 - 인터넷 사이트
 - 유튜브 채널명
 - 특정 서비스

15.13.2. 분석 대상

- IT 서비스 용어

- 애플리케이션: 인스타그램, 트위터, 카카오톡, 리브메이트, 페이스북, 페이스북
- 인터넷 사이트명: 네이버, 다음, 구글, 야후, 라이코스, 네티앙, 아이팝콘, 네이버TV, 도그티비, 고양이 라서 다행이야, 세모지, 수미사, 수제나라, 오늘의 유머, 올레닷컴, 아프리카 TV, 트위치
- 유튜브 채널명: 씨티비, 땅고부부, 딘디, 빼약스핏, 오딜
- 특정 서비스: 와이브로 서비스, DMB 서비스, 위성 DMB, 엑스박스라이브, IPTV, 아날로그 방송, 디지털 방송

15.13.3. 예문

- IT 서비스 용어
 - 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.
 - 대표적인 인터넷 포털사이트로는 <네이버: TMI_SERVICE>, <다음: TMI_SERVICE> 등이 있다.
 - 며칠 전 <고양이라서 다행이야: TMI_SERVICE> 카페에 들어갔다가 재밌는 사진을 발견했다.
 - <땅고부부: TMI_SERVICE>와 함께 재미있는 운동을 해 보세요.
 - SK텔레콤은 2018년을 끝으로 <와이브로 서비스: TMI_SERVICE>를 종료했다.
 - 스마트폰이 대중화됨에 따라 <위성DMB: TMI_SERVICE>는 역사 속으로 사라졌다.

15.13.4. 주의 사항

- (태깅 대상 - 특정 서비스) 엑스박스라이브와 같이 서비스만을 대상으로 하며, 다른 세부 유형과 혼동되는 경우에는 다른 세부 유형에 우선권을 부여한다.
 - 엑스박스라이브, PS Plus, 유튜브 프리미엄, 유플레이 플러스, 오리진 엑세스 등
- (태깅 대상- 줄임말) '카톡, 페북, 인스타' 등 특정 애플리케이션의 줄임말도 TMI_SERVICE로 태깅한다.
 - 세금 상담은 카톡으로 하세요. → 세금 상담은 <카톡: TMI_SERVICE>으로 하세요.
- (태깅 대상 제외 - SNS) 온라인 서비스를 의미하는 보통명사 'SNS'는 태깅하지 않는다.
 - SNS는 페이스북과 트위터와 같은 다양한 서비스를 말한다. → SNS는 <페이스북: TMI_SERVICE>과 <트위터: TMI_SERVICE>와 같은 다양한 서비스를 말한다.
- (태깅 대상 제외 - 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.
- (태깅 대상 제외 - OO스타그램) '인형스타그램, 먹스타그램' 등과 같은 복합어 'OO스타그램'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.

15.14. TMI_PROJECT

15.14.1. 정의 및 설명

- 프로젝트 명칭
 - IT 프로젝트
 - 국가 프로젝트
 - 회사 프로젝트
 - 기타 프로젝트

15.14.2. 분석 대상

- 프로젝트 명칭
 - 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
 - 기타 프로젝트: 머큐리 계획, 가오편 프로젝트, 아폴로 미션

15.14.3. 예문

- 프로젝트 명칭
 - 문재인 대통령은 <4대강 사업: TMI_PROJECT> 감사 필요성을 제기하고, 사업 전반의 과정을 밝히는데 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 대상 - OO 사업) '편찬 사업'과 같이 '사업'을 의미할 때는 TMI_PROJECT로 태깅한다.
 - '겨레말 큰사전' 편찬 사업이 4년 만에 좌초 위기를 맞고 있다. → <'겨레말 큰사전' 편찬 사업: TMI_PROJECT>이 4년 만에 좌초 위기를 맞고 있다.
- (태깅 대상 제외 - 공익캠페인) '자연 보호 캠페인'과 같은 '공익캠페인'은 TMI_PROJECT로 태깅하지 않는다.
 - 일도2동 적십자봉사회 10주년 기념 자연 보호 캠페인 전개

15.15. TMIG_GENRE

15.15.1. 정의 및 설명

- 게임 장르
 - 플레이 방식
 - 플레이 환경
 - 기타 게임 장르

15.15.2. 분석 대상

- 게임 장르
 - 플레이 방식: RPG, AOS, FPS, 롤플레이팅 게임, 어드벤처 게임, 아케이드 게임, 시뮬레이션 게임, 전략 게임, 스포츠 게임, 퍼즐 게임, 리듬 게임, 레이싱 게임, 공포 게임, 슈팅 게임, 추리 게임, 액션 게임, 판타지 게임, 무협 게임, 방치형 게임
 - 플레이 환경: PC 게임, VR 게임, 컴퓨터 게임, 온라인 게임, 모바일 게임, 콘솔 게임, 비디오 게임, 플래시 게임, 멀티 플랫폼 게임, 클라우드 게임
- 기타 게임 장르: 사행성 게임, 인디 게임, 성인 게임, 가차 게임, 글로벌 게임, 한국사 컴퓨터 게임

15.15.3. 예문

- 게임 장르
 - 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 제외 – 단순 게임명) 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

- 레저·스포츠 규칙

15.16.2. 분석 대상

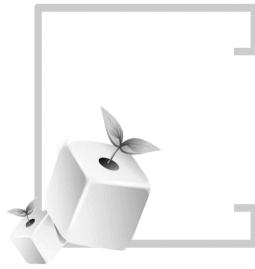
- 레저·스포츠 기술
 - 축구 기술 용어: 발리슛, 라보나킥, 세트피스, 오버헤드킥, 인프런트킥, 토킥, 간접프리킥, 스루패스, 크로스 트레이닝, 드리블, 사포, 스로인, 티키타카, 도움주기
 - 농구 기술 용어: 버저비터, 레이업슛, 리바운드, 차징, 박스 아웃, 엘리웁, 크로스오버, 백다운, 스텝백
 - 야구 기술 용어: 스트라이크, 병살, 더블플레이, 안타, 홈런, 견제구, 고의사구, 번트, 벤치클리어링, 보살, 더블헤더, 사구, 사사구, 희생플라이, 노히트노런, 배트 플립, 투런포, 1점 홈런, 만루 홈런
- 레저·스포츠 규칙
 - 축구 규칙 용어: 오프사이드, 캐링, 핸들링
 - 농구 규칙 용어: 실린더 룰, 라인크로스, 더블 드리블
 - 야구 규칙 용어: 보크, 몰수게임, 콜드게임

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.
- 레저·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 대상 - 수량 관련 스포츠 용어) '볼넷, 삼진'과 같은 야구 관련 용어나 '한판'과 같은 유도 용어는 QT_SPORTS가 아닌 TM_SPORTS로 태깅한다. 이외에 수량 관련 스포츠 용어는 QT_SPORTS로 태깅한다.
 - 기아 김민식은 안타를 치고 1루로 향했다. → 기아 김민식은 안타를 치고 <1루: TM_SPORTS>로 향했다.
 - 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.
 - 심규범은 <1이닝: QT_SPORTS> <1볼넷: QT_SPORTS> 무실점을 기록했다.
- (태깅 단위 - 방향성 정보+야구 용어) 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.



부록

개체 연결 말뭉치 구축 지침



개체 연결 말뭉치 구축 지침

ver. 1.6

2022.03.04.

개체 연결 말뭉치 구축 지침

ver.1.6

이 문서는 개체 연결 말뭉치 구축 지침으로서 태그가 부착된 개체명과 지식베이스(위키피디아)를 연결하는 방안을 기술하는 것을 목적으로 한다. 이 지침은 ‘국립국어원’의 <2020년 개체명 분석 말뭉치 구축 지침>을 기반으로 하며, 15개의 개체명 대분류 태그 세트 중에서 PS(PERSON), AF(ARTIFACTS), OGG(ORGANIZATION), LC(LOCATION), DT(DATE) 등 5개의 대분류에 속하는 개체명을 위키피디아 문서와 연결하는 방안을 기술한다.

1. 개체명과 지식베이스(위키피디아)

1.1. 개체명

개체명은 ‘다른 것들과 구별하여 부르기 위해 이름을 붙인 실제 세계에 존재하는 개체들의 명칭’이다. ‘국립국어원’의 <2020년 개체명 분석 말뭉치 구축 지침>에서는 개체명의 의미 범주를 구분하여 태깅하는 방안을 제시하고 있는데, 개체명 분류 태그 세트는 15개의 대분류와 145개의 소분류, 7개의 세분류, 총 150개의 세부분류로 구성되어 있다. 개체명 태깅 대상에는 대부분의 고유명과 일부의 보통명사가 포함되어 있다. 개체명 태깅 단위는 대체로 단어이나, 하나의 개체를 가리키는 경우 구가 태깅 단위가 되기도 한다.

개체명 태깅은 대상 문서의 맥락에서 특정 개체명이 어떤 의미로 쓰였는지 판별하여 맥락의 의미에 부합하는 개체명 태그를 부착하는 방법으로 한다. 따라서 어떤 개체명이 맥락과 관계없이 절대적인 하나의 태그를 갖는 것이 아니라, 맥락에 따라 서로 다른 태그를 갖기도 한다. 예를 들어 ‘청와대’의 경우 대상 문서에서 ‘행정 기관’의 의미로 사용되었을 경우 OGG_POLITICS로 태깅하고, ‘관저’의 의미로 사용되었을 경우 AF_BUILDING으로 태깅한다. 동일한 개체명이라고 하더라도 의미 범주에 따라 다른 태그를 부착하므로, 언어 표현의 다의성이 서로 다른 태그를 통해 표시된다.

1.2. 위키피디아

위키피디아는 다중 언어 웹 기반 자유 콘텐츠 백과사전으로, 인터넷 사용자라면 누구나 새로운 항목을 작성하거나 수정할 수 있다. 위키피디아는 새로운 항목이 무분별하게 등재되는 것을 막기 위해 ‘일정 기간 세상으로부터 중요한 관심을 받는 주제’, ‘신뢰할 만한 2차 출처가 다수 존재하는

주제'의 경우 등재될 수 있다는 기준을 두고 있다. 어떤 주제에 대해 신뢰할 수 있는 출처가 없거나 중요하게 다루어지지 않는다면, 그 주제는 독립된 문서로 남겨지지 않고 삭제되거나 다른 문서에 병합된다.

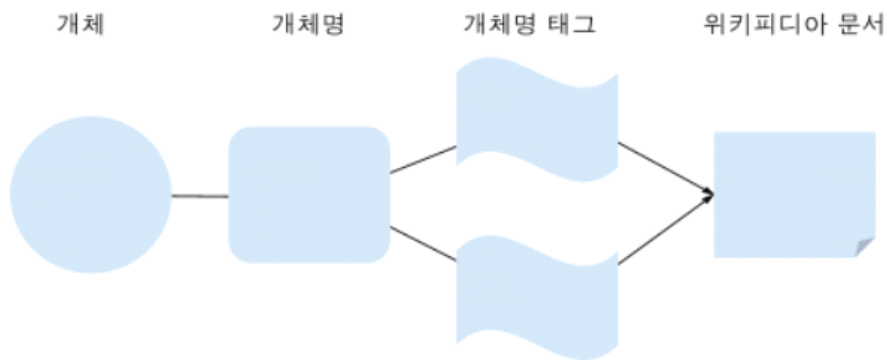
위와 같은 기준을 충족하는 주제는 별도의 문서로 작성할 수 있는데, 문서의 제목은 사용자가 의미의 혼동 없이 쉽게 알아볼 수 있고 다른 문서에서 링크를 걸기 쉬운 것을 선택하도록 하고 있다. 그런데 동음이의어와 같이 의미 혹은 내용은 다르더라도 동일한 제목이 필요한 경우가 있다. 위키피디아는 하나의 제목에 하나의 문서만을 작성할 수 있으므로, 이러한 경우에는 제목 뒤에 한 칸을 띄우고 괄호 안에 알맞은 설명을 써서 동음이의 문서가 쉽게 구분되도록 한다.

위키피디아는 언어사전이 아니므로 단순히 단어의 사전적 정의만을 담지 않고 정의 이상의 다양한 내용을 폭넓게 담을 수 있도록 확장해야 한다고 하고 있으며, 단어 자체가 독립된 문서의 주제가 되더라도 정의 이상의 내용이 포함되거나 사용 사례를 제시하도록 하고 있다. 이러한 지침에 따라 한 단어의 다양한 의미는 별도의 문서가 아닌 하나의 문서에 포함되는 경우가 많다. 그리하여 개체명 태깅 시에는 서로 다른 태그가 부착된 개체명이 개체 연결 시에는 하나의 문서에 연결될 수도 있다.

1.3. 개체명과 지식베이스(위키피디아) 문서 간의 관계 유형

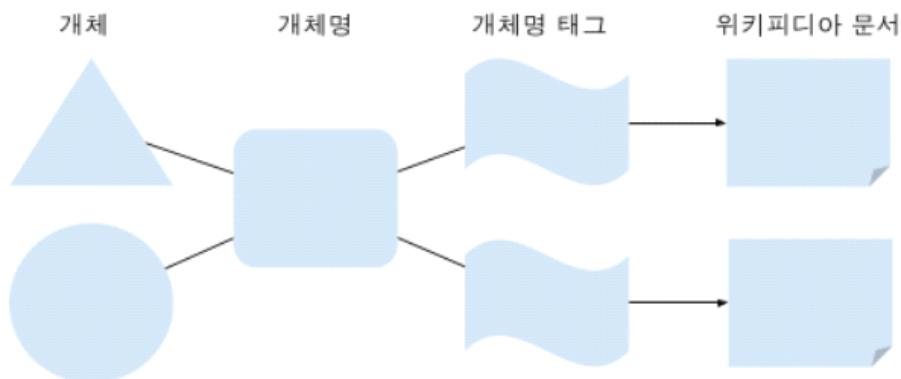
개체명과 위키피디아 문서를 올바르게 연결하기 위해서는 개체(실제 세계의 대상), 개체명(언어 표현)과 개체명 태그(범주적 성격 표지), 그리고 위키피디아 문서(백과사전적 정보) 간의 관계를 분명히 인식할 필요가 있다. 여기에서는 이들 간의 관계 유형을 도식으로 표현하고, 각각의 사례를 제시할 것이다. 1.3에서 기술하는 것은 개체명과 지식베이스(위키피디아) 문서 간의 관계 유형일 뿐 이들 모두를 연결해야 한다는 것은 아니다. 개체명과 지식베이스(위키피디아) 문서 연결의 원칙은 2장에서 상세히 다룬다.

(1) 다의어

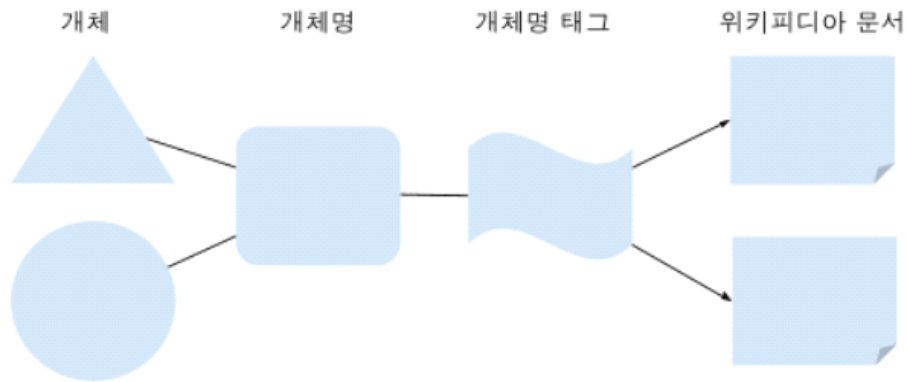


이에 해당하는 사례로 ‘청와대’가 있다. ‘청와대’는 ‘기관’의 의미를 갖기도 하고 ‘관저’의 의미를 갖기도 하는데, 개체명 태그는 ‘기관’일 경우 ‘OGG_POLITICS’이고 ‘관저’일 경우 ‘AF_BUILDING’이다. 위키피디아는 이러한 다의어의 경우에는 별도의 문서를 두지 않고 하나의 문서에서 다양한 의미를 기술한다.

(2) 동음이의어

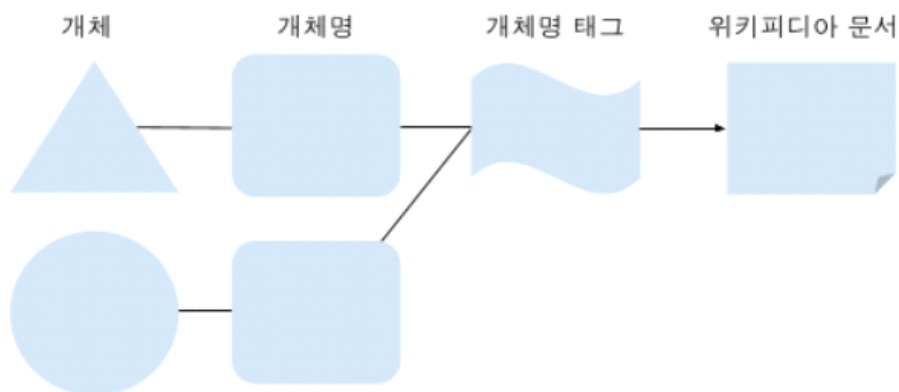


위의 도식과 부합하는 사례로 ‘가을방학’을 들 수 있다. ‘가을방학’은 ‘가을에 일정 기간 동안 수업을 쉬는 일. 또는 그 기간’을 가리키기도 하지만 ‘가을방학’이라는 음악 그룹을 지칭하기도 한다. ‘가을방학’이라는 언어 표현이 지시하는 개체는 ‘기간’과 ‘음악 그룹’으로 각각 다르며, 개체명 태그 또한 ‘DT_DURATION’과 ‘PS_NAME’으로 상이하다. 이러한 ‘가을방학’에 대해 위키피디아에서는 기간을 주제로 하는 ‘가을 방학’ 문서와 음악 그룹을 주제로 하는 ‘가을방학’ 문서를 제공하고 있다.



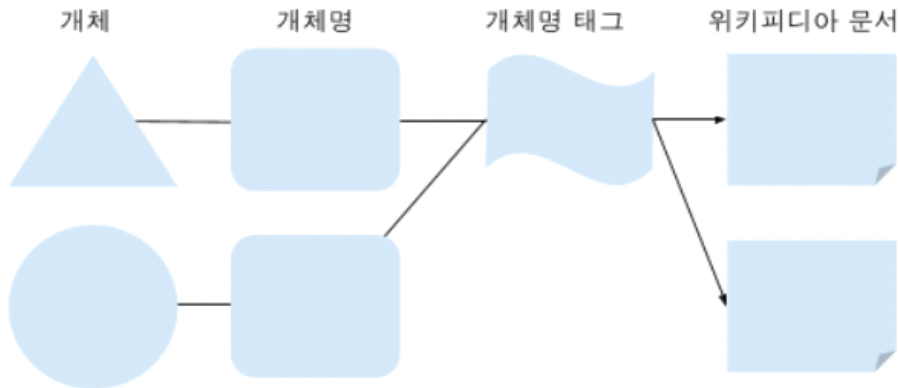
이와 달리 동음이의어이면서도 개체명 태그는 같은 경우도 있다. 인명의 경우 지시하는 개체는 다르지만 개체명은 같으며, 개체명 태그 또한 PS_NAME으로 동일하다. 예를 들어 ‘김정은’이라는 이름을 가진 사람은 무수히 많다. 유명 인사만 하더라도 ‘가수 김정은’, ‘배우 김정은’, ‘농구선수 김정은’ 등이 존재한다. 이러한 ‘김정은’은 지시하는 개체는 다르지만 개체명 태그는 ‘PS_NAME’으로 동일하다. 위키피디아는 이러한 인물명에 대해 ‘김정은 (배우)’와 같이 직업이나 출생연도와 같은 관련 정보를 괄호 안에 표기함으로써 중복을 피하고 있다.

(3) 상위어-하위어



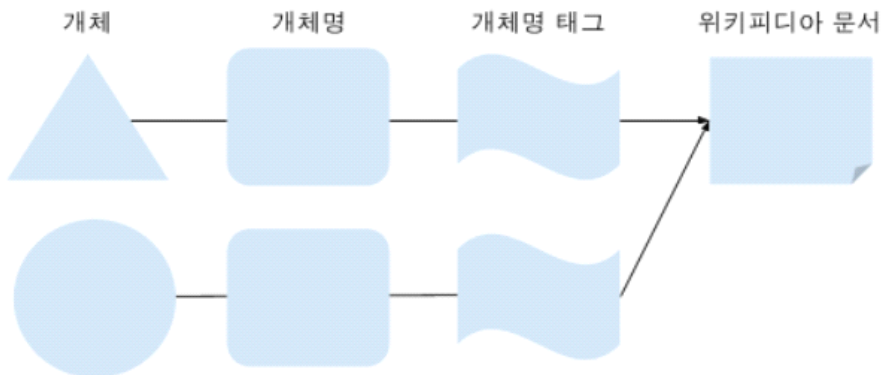
상위 기업과 하위 지점/지부가 이에 해당된다. ‘스타벅스’라는 기업은 ‘스타벅스 인사점’, ‘스타벅스 춘천구봉산R점’ 등 다양한 지점을 보유하고 있다. 개체명 태깅 시 상위 기업과 하위 지점을 하나의 단위로 하여 ‘OGG_FOOD’로 태깅하도록 되어 있어, 상위 기업만 쓰인 ‘스타벅스’와 하위 지

점도 쓰인 ‘스타벅스 인사점’은 모두 ‘OGG_FOOD’로 태깅된다. 그러나 위키피디아에는 상위 기업의 경우 별도의 문서로 존재하나, 하위 지점은 특별한 경우가 아니면 별도의 문서로 존재하지 않고 상위 기업 문서에서 관련 정보로서 제시된다.



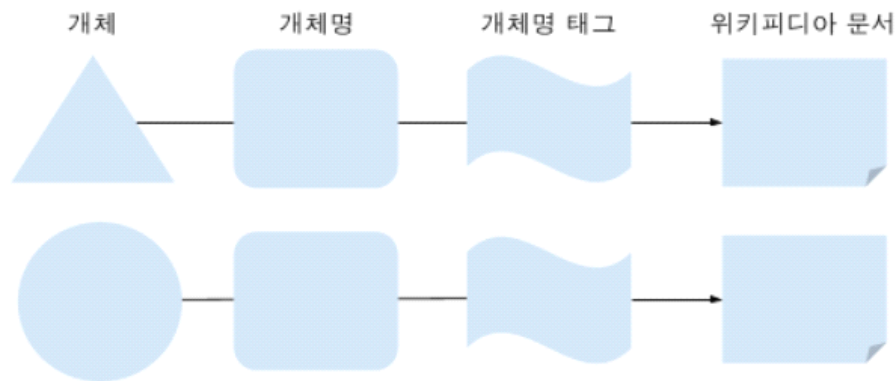
위에서 언급한 바와 같이, 상위 기업과 하위 지점/지부 관계임에도 불구하고 위키피디아에 문서가 별도로 존재하는 경우도 있다. 위키피디아는 ‘웹 기반 자유 콘텐츠 백과사전’이므로 인터넷 사용자라면 누구나 위키백과 항목을 작성하거나 수정할 수 있어, 기준이나 체계가 절대적이지 않고 변동성이 크다. 따라서 사용자가 ‘일정 기간 세상으로부터 중요한 관심을 받는 주제’이면서 ‘신뢰할 만한 2차 출처가 다수 존재하는 주제’라고 판단한 것이면 하위 지점/지부도 별도의 항목으로 존재할 수 있는 것이다. 이러한 예로 ‘신세계백화점 센텀시티’를 들 수 있다. ‘신세계백화점 센텀시티’는 ‘신세계백화점’의 여러 지점 중 하나이므로, ‘스타벅스’-‘스타벅스 인사점’과 같은 관계이다. 그러나 ‘신세계백화점 센텀시티’는 별도의 문서로 존재하는데, 이는 ‘세계 최대 규모의 백화점’으로 기네스북에 등재되었기 때문인 것으로 보인다. 신세계백화점의 다른 지점들은 별도의 문서로 존재하지 않고, ‘신세계 (백화점)’ 문서에서 관련 정보로서 제시된다.

(4) 전체어-부분어



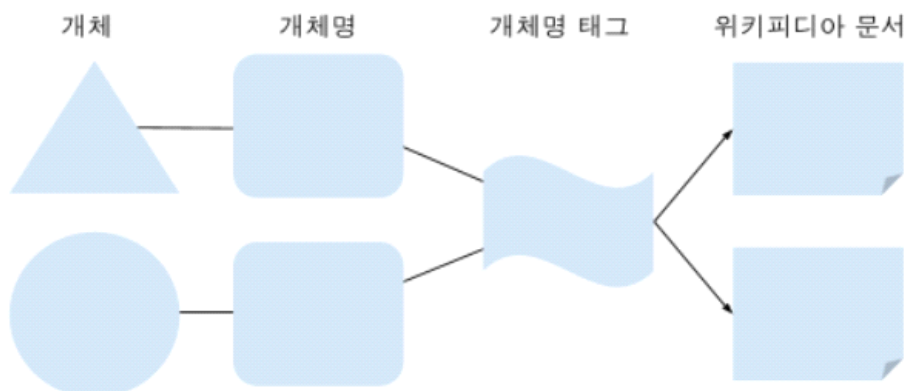
일반적으로 가수가 발매하는 노래는 앨범에 수록되어 있는 경우가 많은데 ‘데칼코마니’라는 노래

는 마마무의 네 번째 미니 앨범 《MEMORY》에 수록되어 있다. 노래 ‘데칼코마니’의 개체명 태그는 AFA_MUSIC, 앨범 《MEMORY》의 개체명 태그는 ‘AFW_OTHER_PRODUCTS’로 서로 다르다. 그러나 위키피디아는 앞에서 설명한 기준에 의해, 특정 노래 문서가 따로 존재하지 않고 앨범 문서에서 관련 정보로서 다루는 경우가 많다. 단, ‘일정 기간 세상으로부터 중요한 관심을 받은 노래’의 경우에는 별도의 문서로 존재한다. 그러한 경우는 아래와 같은 도식으로 표현된다.

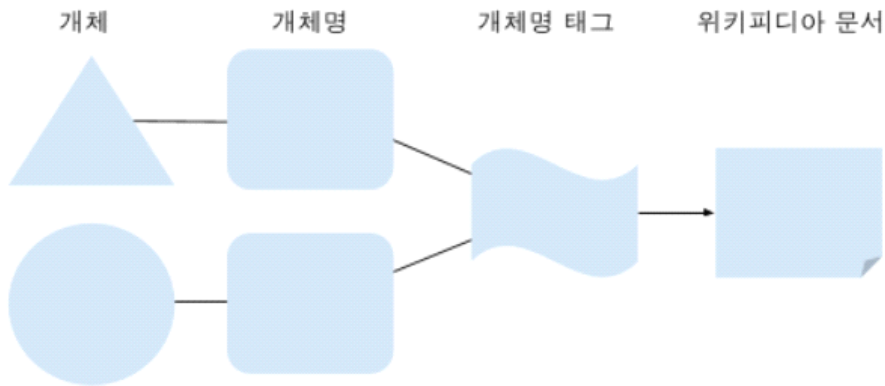


이에 해당하는 사례로 방탄소년단의 앨범 《YOU NEVER WALK ALONE》과 그 타이틀곡 ‘봄날’이 있다. ‘봄날’은 《YOU NEVER WALK ALONE》에 수록되어 있는데, 《YOU NEVER WALK ALONE》도 별도의 항목으로 문서가 존재하고, ‘봄날’ 또한 그렇다. ‘봄날’은 공개 후 26시간 38분 만에 1천만 뷰를 돌파하며 K-pop 그룹 사상 최단 기록을 세우는 등의 특징적인 내용이 있어 별도의 항목으로 다루고 있는 것으로 보인다.

전체어-부분어 중에 개체명 태그가 동일한 경우도 있다. 음악 그룹과 음악 그룹에 속한 멤버가 그 예이다. <2020년 개체명 분석 말뭉치 구축 지침>에 따라 음악 그룹명과 인명은 모두 PS_NAME으로 태깅되는데, 위키피디아에서는 음악 그룹과 그 멤버를 별도의 문서로서 다루기도 하고, 음악 그룹 문서에서 관련 정보로서 소속 멤버들의 정보를 제공하기도 한다. 먼저, 음악 그룹과 그 멤버를 별도의 문서로 제공하는 경우는 아래와 같은 도식으로 표현된다.

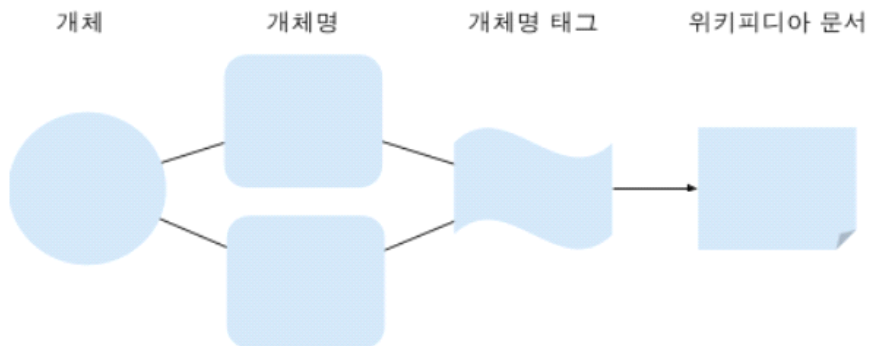


그룹 ‘마마무’의 멤버는 ‘솔라’, ‘문별’, ‘휘인’, ‘화사’ 총 4명으로 이들에 대한 정보는 위키피디아에 별도의 문서로 등록되어 있다. 전체어 ‘마마무’와 부분어 ‘솔라’, ‘문별’, ‘휘인’, ‘화사’가 모두 별도의 문서로 존재하는 것이다.



음악 그룹과 그 멤버를 별도의 문서로 제공하지 않고, 음악 그룹 문서에서 관련 정보로서 소속 멤버들의 정보를 제공하는 경우는 위와 같은 도식으로 표현된다. 그룹 ‘온앤오프’는 대한민국의 6인조 보이 그룹인데 위키피디아에 ‘온앤오프’ 문서는 존재하나, 그 구성원인 ‘효진’, ‘제이어스’, ‘이션’, ‘와이엇’, ‘MK’, ‘유’에 대한 문서는 존재하지 않고, ‘온앤오프’ 문서 안에 이들의 정보가 포함되어 있다.

(5) 동일지시적 표현



어떠한 대상을 부르는 언어 표현이 둘 이상인 경우가 있다. 줄임말, 별칭, 일상적 표현 등의 사용으로 인해 하나의 대상을 지칭하는 언어 표현이 여럿 존재하게 된다. 예를 들어 방탄소년단의 리더를 지칭하는 말은 ‘김남준’, ‘랩몬스터’, ‘랩몬’, ‘알엠’, ‘RM’ 등으로 다양하다. 이러한 동일한 개체를 지시하는 다양한 명칭은 언어 표현은 다를지라도 개체명 태그는 동일하다. 위키피디아에서는 어떤 대상에 대한 문서를 제공할 때 그 제목으로 ‘정식 명칭’을 사용하도록 하고 있어, 각각의 명칭에 대한 문서가 별도로 존재하지 않으며, 해당 대상을 지칭하는 다른 표현들은 관련 정보로서 문서 안에서 다루어진다.

위의 도식을 통해 특정 개체에 대한 개체명과 개체명 태그, 위키피디아 문서는 일대일로 대응될 수도 있고, 일대다로 대응될 수도 있다는 것을 알 수 있다. 이러한 특성을 고려하여 개체명과 위키피디아 문서를 연결할 수 있어야 할 것이다.

2. 개체명과 지식베이스(위키피디아) 문서 연결의 원칙

2.1. 대원칙

2.1.1. 연결 대상

(1) 개체명과 위키피디아 문서 연결의 일차적 기준은 제목이다. 개체명과 동일한 제목의 위키피디아 문서를 찾아 연결한다.

○ 예제

■ 기생충

■ <기생충: AFA_VIDEO>은 봉준호 감독의 작품 중에서 가장 흥행에 성공한 작품으로 꼽힌다.

■ (개체명) 기생충: AFA_VIDEO → (위키피디아) 기생충 (영화)

개체명	위키피디아
기생충 AFA_VIDEO	<p>기생충 (영화)</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>《기생충》(寄生蟲, Parasite)은 2019년 5월 30일에 개봉한 대한민국의 블랙 코미디 서스펜스 영화이다. 봉준호 감독의 일곱 번째 장편 영화로, 한진영과 공동 각본을 썼다. 이 영화는 송강호, 이선균, 조여정, 장혜진, 이정은, 최우식, 박소담과 박영훈 등이 출연한다.</p> <p>대한민국에서 천만 명 이상의 관객을 동원하여 흥행하였으며, 제72회 칸 영화제 공식경쟁 부문에 초청되어 5월 21일 칸에 위치한 비에르 국립에서 공식 스크리닝이 진행되었으며, 황금종려상 수상작인 2013년 《가을 피아노》, 《봄 여름 가을 겨울》 이래 만장일치로 수상한 최초의 대한민국 영화가 되었다. 2019년 5월 30일에 CJ 엔터테인먼트 배급으로 대한민국에서 개봉되었다. 이 영화는 비평가들로부터 폭 넓은 찬사를 받았으며, 2019년 최고의 영화 중 다수로 선정되었다.^[P]이 지금까지 전세계에서 2억 5,755만 달러의 흥행 수익을 벌어들여 가장 높은 수익을 올린 영화이자 대한민국에서 세 번째로 높은 수익을 올리는 영화가 되었다.^[Q]</p> <p>수많은 수상 목록 중에서도 《기생충》은 제77회 골든 글로브상에서 작품상, 감독상, 외국어 영화상 3개 부문에 후보로 선정되었으며, 한국 영화 최초로 외국어영화상을 수상했다. 미국 배우 조합상(SAG)에서 캐스팅 앙상블상을 수상한 최초의 비영어권 영화가 되었다.^[R] 제73회 영국 아카데미 영화상에서 작품상, 감독상, 각본상과 외국어 영화상 4개 부문에 후보로 지명되어 각본상과 외국어 영화상을 수상하였다. 영국 아카데미 영화상에서 작품상 후보에 오른 것은 한국 최초이다. 제92회 아카데미상에서 작품상, 감독상과 각본상, 국제영화상, 편집상, 미술상 4개 부문에 후보로 지명되었으며, 국제영화상 부문 후보에 오른 최초의 한국 영화가 되었다. 영화는 비평가들 대 큰 호평을 받아 작품상, 국제영화상, 감독상, 각본상의 4개의 상을 수상하는 쾌거를 이뤄내며, 역대 시상식 사상 작품상을 수상한 최초의 비영어 영화가 되었다.</p> 

● 단, 개체명과 위키피디아 문서의 제목이 일치하지 않더라도 동일한 대상(개체)에 대한 문서이면 연결한다.

○ 예제

■ 국제연합

■ 농식품부, <국제연합: OGG_OTHERS> 식량농업기구(FAO) 총회 참석

■ (개체명) 국제연합: OGG_OTHERS → (위키피디아) 유엔

(동일한 대상을 지시하므로 연결함.)

개체명	위키피디아
국제연합 OGG_OTHERS	

- 개체명과 정확히 일치하는 제목의 위키피디아 문서는 존재하지 않으나, ‘N1(과) N2’와 같이 개체명의 언어 표현이 위키피디아 문서의 제목에 완전히 포함되고 지시 대상도 완전히 포함되는 문서가 존재하는 경우, 개체명이 그중 일부만을 지시한다고 하더라도 해당 위키피디아 문서로 연결한다.

○ 예제

- 수호랑
- 역대 올림픽 마스코트 중에서 <수호랑: PS_CHARACTER>이 제일 귀여워.
- (개체명) 수호랑: PS_CHARACTER → (위키피디아) 수호랑과 반다비


개체명	위키피디아
수호랑 PS_CHARACTER	

- (2) 개체명이 지시하는 대상과 위키피디아 문서에서 설명하는 대상이 완전히 일치하는 경우에 한해 연결한다.

○ 예제

- 아이폰 X
- 아이폰 10주년 기념작 <아이폰 X: AFW_OTHER_PRODUCTS>.

■ (개체명) 아이폰 X: AFW_OTHER_PRODUCTS → (위키피디아) 아이폰 X

개체명	위키피디아
<p>아이폰 X</p> <p>AFW_OTHER_PRODUCTS</p>	<p>아이폰 X</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>아이폰 X(iPhone X, 아이폰 텐^(en))은 애플이 설계, 개발한 스마트폰의 하나이다. 2017년 9월 12일에 애플 파크 캠퍼스에 있는 스티브 잡스 시어터에서 애플의 CEO 팀 쿡에 의해 아이폰 8, 아이폰 8 플러스와 함께 발표되었으며 아이폰 X는 아이폰의 10주년을 기념하기 위해서 개발해왔다. X는 로마 숫자 10을 뜻하여 '아이폰 텐'이라고 부른다.</p> <p>애플의 라인업에서 X 모델은 아이폰의 프리미엄 모델에 위치해 있으며, 진보된 기술을 공개하기 위해 고안되었다. 케이스 제조업체들, 홀트 필웨어, 최신 버전의 iOS 11에서 유출된 사람들은 이 장치의 다양한 관점을 드러냈으며 여기에는 베젤이 거의 없는 디자인, 물리적 홈 버튼의 부재, OLED 디스플레이, 개선된 심도 감지, 또 페이스 ID로 불리는 안면 인식 잠금 해제 시스템을 포함하고 있다. 참고로, 최초로 화면 모서리가 둥글어지고, 3차로 곡면 본체이다.^[78]</p> <div> <div> <p>목차 (펼치기)</p> <ol style="list-style-type: none"> 역사 사양 <ol style="list-style-type: none"> 하드웨어 소프트웨어 출시 국가 각주 외부 링크 </div> <div>  <p>아이폰 X</p> <p>Apple Inc.</p> </div> </div>

(3) 개체명 태그와 동일한 속성을 갖는 위키피디아 문서를 연결한다. 예를 들어, 개체명이 AFA_ART_CRAFT로 태깅되어 있다면 예술 작품에 대한 위키피디아 문서를, 개체명이 AFA_VIDEO로 태깅되어 있다면 영화 혹은 드라마에 대한 위키피디아 문서를 찾아 연결한다.

○ 예제

- 별이 빛나는 밤
- 빈센트 반 고흐의 ‘〈별이 빛나는 밤: AFA_ART_CRAFT〉’
- (개체명) 별이 빛나는 밤: AFA_ART_CRAFT → (위키피디아) 별이 빛나는 밤

개체명	위키피디아
<p>별이 빛나는 밤</p> <p>AFA_ART_CRAFT</p>	<p>별이 빛나는 밤</p> <p>위키백과, 우리 모두의 백과사전.</p> <p> 다른 뜻에 대해서는 별이 빛나는 밤 (동음이의) 문서를 참고하십시오.</p> <p>《별이 빛나는 밤》(The Starry Night)은 네덜란드의 화가 빈센트 반 고흐의 가장 널리 알려진 작품이다. 정신병을 앓고 있을 당시의 고흐가 그린 작품이다. 1889년 상 레미의 정신병원에서 나와 기역으로 그린 그림이지. 당시 고흐는 정신적 질환으로 인한 고통을 떠올려 그림 속의 소용돌이로 묘사한다.^[1]</p> <p>빈센트 반 고흐의 대표작 중 하나로 꼽히는 《별이 빛나는 밤》은 그가 고흐와 다른 뒤 자신의 귀를 자른 사건 이후 생애의 요양원에 있을 때 그린 것이다. 그는 별빛 밖으로 내다보이는 방 풍경들 기억과 상상을 결합시켜 그렸는데, 이는 자연에 대한 반 고흐의 내적이고 주관적인 표현을 구현하고 있다.^[2] 고흐에게 밤하늘은 유한함을 표현하는 대상이었고, 이보다 먼저 제작된 작품의 《밤의 카페 테라스》나 《론 강 위로 별이 빛나는 밤》에서도 별이 반짝이는 밤의 장면을 다루었다. 고흐는 《별이 빛나는 밤》의 작업을 마칠 때 그다지 좋게 생각하지 않았다고 한다. 작품이 소개될 당시 미술계의 반응도 변변찮았다. 현재 이 그림은 1941년부터 뉴욕 현대미술관(MAMA)에서 상설 작품으로 전시되고 있다.^[3]</p> <p>연속적이고 동적인 터치로 그려진 하늘은 곧이어지는 두려운 붓놀림으로 사이프러스와 연결되고, 그 아래의 마을은 대조적으로 고요하고 평온한 상태를 보여준다. 교외 집합은 그의 고향인 네덜란드를 연상시킨다. 그는 별빛 밖으로 내다보이는 방 풍경을 상상과 결합시켜 그렸는데, 이는 자연에 대한 반 고흐의 내적이고 주관적인 표현을 구현하고 있다. 수직으로 놓여 별이 땅과 하늘을 연결하는 사이프러스는 전통적으로 무덤이나 애도와 연관된 나무지만, 반 고흐는 지옥을 불길하게 보지 않았다.^[4]</p> <div> <div> <p>별이 빛나는 밤</p>  </div> <div> <p>작가 빈센트 반 고흐</p> <p>연도 1889</p> <p>매체 캔버스에 유채</p> <p>세로 팔만삼무의</p> <p>크기 73.5cm x 92.1cm</p> <p>소장 뉴욕 현대미술관</p> </div> </div>

○ 예제

- 별이 빛나는 밤
- 제16회 부산국제영화제 상영작 ‘〈별이 빛나는 밤: AFA_VIDEO〉’

- (개체명) 별이 빛나는 밤: AFA_VIDEO → (위키피디아) 별이 빛나는 밤 (영화)

개체명	위키피디아
별이 빛나는 밤 AFA_VIDEO	<p>별이 빛나는 밤 (영화)</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>《별이 빛나는 밤》(원제: 스타리 스타리 나이트; 空想, 공식 영문명: <i>Starry Starry Night</i>)는 2011년 제작된 타이완의 영화이다. 대한민국에서는 2011년에 제16회 부산국제영화제에서 상영되어 뉴 커리츠 후보에 올랐고(관유유), CGV 무비플러스에서 수입하여 2012년 4월 5일에 극장 개봉하였다.</p> <p>제작 (출가기)</p> <ol style="list-style-type: none"> 1 캐스트 2 스태프 3 해외 수출 <ul style="list-style-type: none"> 3.1 대한민국 4 외부 링크 <p>별이 빛나는 밤 空想</p> <p>감독: 민수용 출연: 이자오, 민후이인 등 개봉일: 2012년 4월 5일 (대한민국) 국가: 타이완 언어: 중국어</p>

(4) 한국어판 위키피디아 문서를 우선으로 하여 연결하되, 한국어판 위키피디아에 개체명이 지시하는 대상에 대한 문서가 존재하지 않는 경우에는 영어판 위키피디아 문서를 찾아 연결한다.

○ 예제

- 보(Bo)
- 오바마 전 대통령은 8일(현지시간) 밤 자신의 트위터에 "오늘 우리 가족은 진정한 친구이자 충직한 동반자를 잃었다"며 <보: PS_PET>(〈Bo: PS_PET〉)의 죽음을 알렸습니다.
- (개체명) 보: PS_PET / (한국어판 위키피디아) 문서 없음. (연결 X)
- (개체명) 보: PS_PET → (영어판 위키피디아) Bo (dog)

개체명	위키피디아
보(BO) PS_PET	<p>Bo (dog)</p> <p>From Wikipedia, the free encyclopedia</p> <p>Bo (October 9, 2008 – May 8, 2021) was a pet dog of the Obama family, the first family of the United States from 2009 until 2017.^[2] President Barack Obama and his family were given the male Portuguese Water Dog as a gift after months of speculation about the breed and identity of their future pet.^{[2][6]} The final choice was made in part because elder daughter Malia's allergies dictated a need for a hypoallergenic breed. Bo was occasionally called "First Dog".^{[1][2][3]} In August 2013, Bo was joined by Sunny, a female dog of the same breed.^[9]</p> <p>Bo was named by sisters Malia and Sasha after their cousins' cat, First Lady Michelle Obama's father, and as a reference to the R&B musician Bo Diddley. Bo's name was also Barack Obama's initials.</p> <p>Bo</p> 

- 개체명이 영문, 한문 등 다른 나라의 문자로 적혀 있을 경우에도 우선적으로 한국어판 위키피디아에서 한글로 재검색하여 개체명이 지시하는 대상을 주제로 하는 문서를 찾아 연결한다. 한국어판 위키피디아에서 한글로 재검색하였을 때 문서가 존재하지 않는다면 영어판 위키피디아 문서를 찾아 연결한다.

○ 예제

- DAEGU
- 2020 SBS 가요대전 in <DAEGU: LCP_CITY>
- (개체명) DAEGU: LCP_CITY → (한국어판 위키피디아) 대구광역시
- (개체명) DAEGU: LCP_CITY / (영어판 위키피디아) Daegu (연결 X)

개체명	위키피디아
DAEGU LCP_CITY	

(5) 비유적 의미는 고려하지 않고 대상 문서에서 쓰인 개체명의 언어 표현과 일치하는 위키피디아 문서를 연결한다.

○ 예제

- 아프리카
- 지난주에 <아프리카: LCG_CONTINENT> 갔다 왔다. 대구 진짜 덥더라. 대프리카 인정.
- (개체명) 아프리카: LCG_CONTINENT → (위키피디아) 아프리카
- 위의 예에서 ‘아프리카’는 바로 뒤의 ‘대구’를 의미하지만, 언어 표현 그대로 위키피디아의 ‘아프리카’ 문서를 연결한다.

개체명	위키피디아
아프리카 LCG_CONTINENT	

● 즉, 개체명 태깅 표현을 기준으로 위키피디아 문서를 연결한다.

○ 예제

- 자동차
- 지난달에 K5를 새로 샀는데 <자동차: AF_TRANSPORT> 잘 나가더라.
- (개체명) 자동차: AF_TRANSPORT → (위키피디아) 자동차
- 위의 예에서 '자동차'는 앞의 'K5'를 의미하지만, 개체명 태깅 표현을 기준으로 위키피디아의 '자동차' 문서를 연결한다.

개체명	위키피디아
<p>자동차</p> <p>AF_TRANSPORT</p>	

2.1.2. 연결 단위

(1) 개체명 태그를 할당할 때 설정한 태깅 단위와 동일한 단위를 제목으로 하는 위키피디아 문서를 연결한다.

- 위키피디아 문서의 연결 단위는 국립국어원의 <2020년 개체명 분석 말뭉치 구축 지침>의 개체명 태깅 단위를 따른다.
- '최소 단위 태깅 원칙'과 '최장 단위 태깅 원칙'에 따라 개체명의 태깅 단위를 설정하였다면, 해당 단위와 일치하는 제목의 위키피디아 문서를 찾아 연결한다.
- 개체명 설정 단위 전체와 일치하는 제목의 위키피디아 문서가 존재하지 않는다면 연결하지 않고 '없음'에 해당하는 정보를 표시한다.
 - (개체명) <현대카드 뮤직 라이브러리: OGG_LIBRARY> / (위키피디아) NONE (문서 없음)
- 개체명을 구성하는 단어 중 일부만을 제목으로 하는 위키피디아 문서를 연결하지 않는다.
 - (개체명) <2021년 7월 9일: DT_OTHERS> / (위키피디아) 2021년 (연결 X)
- 단, '스타벅스 인사점'과 같이 상위 조직과 하위 지점(지부)으로 구성된 개체명의 경우, 구성 전체(스타벅스 인사점)가 별도의 위키피디아 문서로 존재하지 않는다면 상위 조직 문서(스타벅스)를 연결한다. (이와 관련된 내용은 '2.2. 세부 원칙' 참조)
 - 예제
 - 스타벅스 인사점
 - <스타벅스 인사점: OGG_FOOD>에서 토피넛라떼 한 잔!

● (개체명) 스타벅스 인사점: OGG_FOOD → (위키피디아) 스타벅스

개체명	위키피디아
스타벅스 인사점 OGG_FOOD	

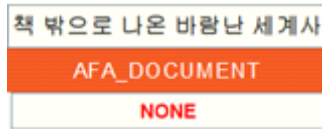
- ‘신세계백화점 센텀시티’, ‘롯데백화점 청량리점’ 등과 같이 하위 지점(지부)에 대한 문서가 위키피디아에 별도로 존재하면 해당 문서를 연결한다.
- 예제
 - 신세계백화점 센텀시티
 - <신세계백화점 센텀시티: OGG_ECONOMY>가 지난 19일부터 3일간 코로나 확진자 발생으로 휴점했다.
 - (개체명) 신세계백화점 센텀시티: OGG_ECONOMY → (위키피디아) 신세계백화점 센텀시티

개체명	위키피디아
신세계백화점 센텀시티 OGG_ECONOMY	

2.1.3. 비연결 항목

- 아래의 항목에 해당하는 경우, 개체명에 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.
 - (1) 개체명과 동일한 제목의 위키피디아 문서가 존재하지 않는 경우
 - 예제
 - 박철규의 ‘〈책 밖으로 나온 바람난 세계사: AFA_DOCUMENT〉’

- (개체명) 책 밖으로 나온 바람난 세계사: AFA_DOCUMENT / (위키피디아) NONE (문서 없음)



(2) 개체명과 동일한 제목의 위키피디아 문서가 존재하나, 의미가 다른 경우 (동음이의어)

○ 예제

- 제주 <천왕사: AF_CULTURAL_ASSET>
- (개체명) 천왕사: AF_CULTURAL_ASSET / (위키피디아) NONE (문서 없음)
- 앞의 '제주'로 보았을 때, '천왕사'는 제주특별자치도에 있는 '천왕사'를 의미하는 것으로 보이나, 위키피디아에는 부여에 있는 '천왕사' 문서만 존재한다. 위키피디아에 개체명과 동일한 제목의 문서가 존재하지만 대상, 의미가 다르므로 연결하지 않는다.

(3) 개체명이 동음이의어인데 대상 문서에서 개체명이 지시하는 대상을 맥락을 통해 특정할 수 없는 경우

○ 예제

- <김정은: PS_NAME> 최고!
- (개체명) 김정은: PS_NAME / (위키피디아) NONE
- 위의 문장에서 '김정은'이 유명 인사인지 일반인인지, 유명 인사라고 하더라도 배우인지 가수인지 농구선수인지 특정할 수 없으므로, 위키피디아에 여러 '김정은' 문서가 존재한다고 하더라도 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

(4) 개체명이 하위어인데, 위키피디아에 하위어 문서가 별도로 존재하지 않고 상위어 문서로 넘어가는 경우(예외적으로 연결하는 사례는 2.2.3. 참조)

○ 예제

- 제주에서 <나룻배: AF_TRANSPORT> 체험하고 왔어요.
- (개체명) 나룻배: AF_TRANSPORT / (위키피디아) NONE
- 위키피디아에서 '나룻배'를 검색하면 상위어인 '선박' 문서로 자동으로 넘어간다. '선박'과 '나룻배'는 상위어-하위어 관계로, 언어 표현과 지시하는 대상이 일치하지 않으므로 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

(5) 개체명이 부분어인데, 위키피디아에 부분어 문서가 별도로 존재하지 않고 전체어 문서로 넘어가는 경우(예외적으로 연결하는 사례는 2.2.4. 참조)

○ 예제

- <어피치: PS_CHARACTER> 보조배터리 너무 귀여워.
- (개체명) 어피치: PS_CHARACTER / (위키피디아) NONE

- 위키피디아에서 ‘어피치’를 검색하면 전체어인 ‘카카오프렌즈 문서로 자동으로 넘어간다. ‘카카오프렌즈’는 카카오가 출시한 ‘라이언, 어피치, 튜브, 콘, 무지, 프로도, 네오, 제이지’ 등의 캐릭터 전체를 지시하는 개념이므로, 개체명 ‘어피치’에 위키피디아 ‘카카오프렌즈’ 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

(6) 개체명 설정 단위 전체와 일치하는 제목의 위키피디아 문서가 존재하지 않는 경우

○ 예제

- 이태원에서 가 볼 만한 곳, <현대카드 뮤직 라이브러리: OGG_LIBRARY>
- (개체명) <현대카드 뮤직 라이브러리: OGG_LIBRARY> / (위키피디아) NONE (문서 없음)
- ‘현대카드’ 문서는 존재하나, ‘현대카드 뮤직 라이브러리’ 전체에 대한 문서가 아니므로 연결하지 않는다.

(7) 개체명의 본유적 의미와 맞지 않는 개체명 태그가 부착되어 있는 경우

- <2020년 개체명 분석 말뭉치 구축 지침>에 따르면 개체명이 대상 문서의 맥락에서 어떤 의미로 사용되었는지에 따라, 맥락적 의미에 맞는 개체명 태그를 부착하도록 되어 있다. 따라서 개체명의 본유적 의미와 맞지 않는 개체명 태그가 부착될 수 있다. 예를 들어, 대상 문서에서 상품명을 기업명으로 대신 지시할 수 있는데, 이때 기업명은 AF로 태깅된다.
- 이러한 경우에 개체명 태깅 표현을 기준으로 위키피디아 문서를 연결하면 개체명 태그와 위키피디아 문서가 맞지 않는 문제가 발생하고, 개체명 태그를 기준으로 위키피디아 문서를 연결하면 개체명 태깅 표현과 위키피디아 문서가 맞지 않는 문제가 발생한다.
- 따라서 개체명의 본유적 의미와 맞지 않는 개체명 태그가 부착되어 있는 경우, 개체명에 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

■ 예제

- 내가 상당히 어려워하는 작곡가 모차르트의 피아노 소나타 16번. 정말 <모차르트: AFA_MUSIC> 너무 어렵다.
- (개체명) <모차르트: AFA_MUSIC> / (위키피디아) NONE (문서 없음)
- ‘모차르트’는 본유적으로 ‘인물’에 해당한다. 그러나 대상 문서의 맥락에서는 ‘모차르트의 피아노 소나타 16번’ 혹은 ‘모차르트의 작품’의 의미로 쓰여 ‘AFA_MUSIC’으로 태깅되어 있다. 이러한 경우, 위에서 설명한 문제가 발생하므로 위키피디아 문서를 연결하지 않는다.

2.2. 개체명의 어휘 관계에 따른 세부 원칙

2.2.1. 개체명이 다의어인 경우

- (1) 개체명이 다의어인데, 하나의 위키피디아 문서 안에 둘 이상의 의미가 모두 기술되어 있는 경우 서로 다른 세분류로 태깅된 개체명에 해당 위키피디아 문서를 각각 연결한다.
- 언어의 다의성으로 인해, 둘 이상의 세분류로 태깅 가능한 개체명들이 있다. 이 사업에서는 그러한 경우 지침에 따라 문맥의 의미에 맞는 하나의 태그를 부착하도록 하고 있다. 즉, 동일한 개체명이 서로 다른 태그를 가질 수 있는 것이다.
 - 그러나 위키피디아는 어떤 개체가 다의성을 갖는다고 하더라도 별도의 문서로 각각의 의미를 기술하지 않고, 하나의 문서에서 다양한 의미들을 모두 기술한다. 예를 들어, '청와대'는 '청와대(기관)', '청와대(관저)'와 같이 문서가 나뉘어 있지 않고, '청와대'라는 하나의 문서에서 '청와대는 대한민국의 대통령이 기거하는 대통령 관저이자 대한민국 헌법이 규정하는 헌법기관으로서의 대통령부와 관계된 행정기관을 말한다.'와 같이 정의에 기관과 관저의 의미를 모두 포함하고 있다.
 - 이와 같이 하나의 문서에서 둘 이상의 의미를 모두 기술하고 있는 경우에는 서로 다른 세분류로 태깅된 개체명에 해당 위키피디아 문서를 각각 연결한다.

○ 예제

- (기관) 청와대
- <청와대: OGG_POLITICS>는 장관 거취에 대한 기존 입장을 고수했다.
- (개체명) 청와대: OGG_POLITICS → (위키피디아) 청와대

개체명	위키피디아
청와대 OGG_POLITICS	

○ 예제

- (관저) 청와대
- <청와대: AF_BUILDING>에는 대한민국 대통령이 공무를 수행하는 '대통령 집무실'과 퇴근 후 기거하는 '대통령 관저'가 있다.
- (개체명) 청와대: AF_BUILDING → (위키피디아) 청와대

개체명	위키피디아
청와대 AF_BUILDING	

(2) 개체명이 다의어인데, 하나의 위키피디아 문서 안에 하나의 의미만 기술되어 있는 경우에도 해당 위키피디아 문서를 서로 다른 세분류로 태깅된 개체명에 각각 연결한다.

- 위키피디아는 일종의 백과사전으로 하나의 주제를 하나의 문서에 담아 해당 대상이 갖는 다양한 의미와 관련 정보들을 제공한다. 따라서 동음이의어가 아니라면 하나의 주제에 둘 이상의 문서가 존재하지 않는다. 다시 말해, 위키피디아에는 하나의 주제에 대한 의미별 문서가 별도로 존재하지 않고 하나의 문서를 통해 다양한 의미, 정보를 제공하는 것이다.
- 따라서 개체명이 둘 이상의 세분류로 태깅될 수 있는데 위키피디아에는 각각의 문서가 따로 존재하지 않고 하나의 문서만 존재하는 경우에도, 해당 위키피디아 문서를 서로 다른 세분류로 태깅된 개체명에 각각 연결한다.

○ 예제

- (기관) 국립중앙박물관
- <국립중앙박물관: OGG_ART>은 이번 달 25일까지 ‘손백 치며 감탄하네 : 김홍도의 풍속도첩’ 전시를 선보일 예정이다.
- (개체명) 국립중앙박물관: OGG_ART → (위키피디아) 국립중앙박물관

개체명	위키피디아
국립중앙박물관 OGG_ART	

○ 예제

- (건물) 국립중앙박물관
- <국립중앙박물관: AF_BUILDING>은 서울시 용산구에 있다.
- (개체명) 국립중앙박물관: AF_BUILDING → (위키피디아) 국립중앙박물관

개체명	위키피디아
국립중앙박물관 AF_BUILDING	<p>국립중앙박물관 위키백과, 우리 모두의 백과사전.</p> <p>국립중앙박물관(國立中央博物館, 영어: National Museum of Korea)은 대한민국 서울특별시 용산구 서빙고로 117번지에 위치한 국립박물관이다. 1932년 7월 19일 발족하였으며, 현재의 박물관은 2005년 10월 28일 건축 개관한 것이다.</p> <p>본관은 통관과 서관으로 나누어져 있으며, 지하 1층 지상 6층으로 이루어져 있다. 전시 공간은 3개 층으로 나누어져 있다. 길이 404미터, 최고 높이 43.08미터의 건물로 세계에서 4번째로 큰 박물관이다. 이 어린이 박물관과 대입 전시장이 별도로 갖추어져 있다. 1월 1일은 휴관이며, 2008년 5월 1일부터 상설전시관, 어린이박물관의 관람료는 무료가 되었다. 부설국립 공예관은 여러 가지 공예품이며 전통공예품과 현대공예품 등의 부설 시설도 있다. 2017년 11월 1일부터는 문화체육관광부가 보안물 이유로 박물관 입구 앞에 보안검색대(일명 x-ray검색대)가 설치되어 감사를 받고 관람할 수 있게 되어 있다.</p> 

2.2.2. 개체명이 동음이의어를 가지는 경우

- (1) 개체명이 동음이의어를 가지는 경우, 즉 개체명이 지시하는 대상이 여럿 존재하는 경우에는 개체명과 동일한 제목을 지닌 위키피디아 문서 중에서 괄호 안의 추가 정보와 본문의 내용으로 보았을 때 개체명이 지시하는 대상과 일치한다고 판단되는 문서를 찾아 연결한다.

○ 예제

- (신화적 인물) 토르
- <토르: PS_NAME>는 북유럽 신화에 나오는 천둥의 신이다.
- (개체명) 토르: PS_NAME → (위키피디아) 토르

개체명	위키피디아
토르 PS_NAME	<p>토르 위키백과, 우리 모두의 백과사전.</p> <p>다른 문에 대해서는 토르 (동음이의) 문서를 참고하십시오.</p> <p>토르(고대 노르드어: Þor; Old Norse: Þorr)는 노르드 신화의雷神 중 양자를 든雷神, 그와 연관되는 개념으로는 천둥, 번개, 폭풍, 참나무, 철학, 인류의 보호, 청회, 치료, 생산물 등이 있다. 영글로색슨어로는 투노르(Þunor), 고대교회슬라브어로는 도나르(donar)라고 하였다.</p> <p>토르는 로마의 제우스나 정제 때부터 안락 다이몬기를 거쳐, 스칸디나비아의 기독교화가 이루어지기 시작한 바이킹 시대에 이르기까지 게르만 민족들에게 널리 숭배받았다. 현대까지도 토르에 관한 지역 민족은 게르만 문화의 영향을 받은 지역에서 잔존하였다. 토르의 이름은 여러 지역에 남아 있으며, 독일어를 의미하는 "Thursday" 역시 Thor's day로 토르에 그 어원을 두고 있다.</p> <p>아이슬란드 문헌에 잘 보존되어 있는 노르드 신화에서 토르에 관한 많은 이야기와 정보를 얻을 수 있다. 이 문헌에서 토르는 적어도 열네 개의 이름을 가지고 있다. 광탈의 여신 시프의 남편이며 여왕 요툰 아른사사가 애인이다. 또한 대개 눈매가 더러워 험악한 것으로 묘사되며, 해리카락과 수많은 붉은 색이 라고 한다.^[1]</p> <p>토르는 지프와의 사이에서 여신 (발키리라고도 함) 스루드를 낳았고, 아른사사와의 사이에서 다그니를 낳았으며, 이들이 알려쳐 있지 않은 누군가와와의 사이에서 모리를 낳았다. 그리고 그는 올론의 외뿔아버지이다. 같은 문헌에서 토르는 오딘과 대지의 여신 프린간의 아들이라고 하는데, (하르바르드 음모시)를 비롯한 문헌에서는 오딘과 싸우는 모습이 보이기도 해서 다소 의문이 있다.</p> <p>토르는 두 명의 하인 살피와 로스크비를 거느리고, 두 마리 열소 황금리프니르와 황금노스(잡어먹이도 재생한다)가 모는 천차에 타고 다니며, 세 채의 저택(발키리니르, 스루드헤임, 스루드방)을 소유하고 있다. 토르는 산도 부을 수 있는 방치 물니르를 들고, 하리의 메김요르드를 차고, 무쇠 창갑 아틀그레 이르를 걸으며, 그리드르를 비롯한 지왕어를 소유하고 있다. 토르가 그의 작품을 가져왔어 때려죽인 이야기, 괴물 벤 요르문간드와 싸운 이야기, 라그나로크에서 결국 동귀여전하게 되는 이야기가 토르 신화로 전해지고 있다.</p> 

○ 예제

- (가공의 인물) 토르
- 어벤져스 엔드게임에서 <토르: PS_CHARACTER>가 제일 좋았어.
- (개체명) 토르: PS_CHARACTER → (위키피디아) 토르 (마블 코믹스)

개체명	위키피디아
<p>토르</p> <p>PS_CHARACTER</p>	<p>토르 (마블 코믹스)</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>토르 오딘손(Thor Odinson)은 마블 코믹스의 슈퍼히어로이다. 북유럽 신화에서 동명의 신 토르를 기반으로 하고 있다. 곧 아스가르드의 왕이 될 토르는 아스가르드를 공격한 요툰헤임에게 분노하게 되고, 전쟁을 일으킨다. 다행히 아버지 오딘이 토르를 막아 전쟁을 중단하게 되고, 토르는 지구로 추방 당하게 된다. 토르 이름 뒤에 붙은 오딘손은 오딘의 아들이라는 뜻이다.</p> <p>담당 배우 [편집]</p> <p>영화 [편집]</p> <ul style="list-style-type: none"> 마블 시네마틱 유니버스 - 크레스 험스워스 <div> <div>토르 오딘손</div> <div>Thor Odinson</div> <div>《마블 코믹스》 등장인물</div> <div>첫 등장</div> <div>마스타리코의 여왕 제8호 (1962년 8월)</div> <div>배우</div> <div>크레스 험스워스 (MCU)</div> <div>정보</div> <div>다른 이름</div> <div>천둥의 신, 번개의 신, 사촌</div> <div>종족</div> <div>아스가르드인</div> <div>성별</div> <div>남성</div> </div>

- (2) 개체명과 동일한 제목을 지닌 위키피디아 문서 중에서 괄호 안의 추가 정보나 본문의 내용이 개체명이 지시하는 대상과 일치하는 문서가 없는 경우 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

2.2.3. 개체명이 상위어를 가지는 경우

- (1) 개체명이 하위어인데, 위키피디아에서 해당 개체명을 검색하였을 때 상위어 문서로 자동으로 넘어가는 경우, 연결하지 않고 ‘없음’에 해당하는 정보를 표시하는 것을 원칙으로 한다.
- (2) 단, 다수의 지점 혹은 지부를 보유하고 있는 기업이나 기관의 경우, 위키피디아에 상위 조직과 하위 조직의 일부에 대한 문서만 있거나 상위 조직 문서만 있고 하위 조직 문서는 없을 수 있다. 이러한 하위 조직-상위 조직 관계를 갖는 개체명은 예외적으로 다음과 같이 연결한다.
- 개체명이 상위 조직을 의미하는 경우 상위 조직 문서를 연결한다.

○ 예제

- (상위 조직) 이마트
- <이마트: OGG_ECONOMY>는 자연주의, 피코크 등의 다양한 브랜드를 자체적으로 개발하여 판매하고 있다.
- (개체명) 이마트: OGG_ECONOMY → (위키피디아) 이마트

개체명	위키피디아
<p>이마트</p> <p>OGG_ECONOMY</p>	<p>이마트</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>이마트(영어: emart)는 신세계가 1993년에 설립한 대형 유통점인데 2011년 별도 법인으로^{[*]} 분할됐다. 2020년 5월 28일 기준 대한민국, 베트남, 몽골 등지에서 총 497개의 점포를 운영 중이다.</p>

- 개체명이 하위 조직을 의미하는데 해당 하위 조직이 별도의 문서로 존재하는 경우 해당 하

위 조직 문서를 연결한다.

○ 예제

- (하위 조직) 롯데백화점 청량리점
- <롯데백화점 청량리점: OGG_ECONOMY>이 6월 24일부터 동행세일을 시작한다.
- (개체명) 롯데백화점 청량리점: OGG_ECONOMY → (위키피디아) 롯데백화점 청량리점

개체명	위키피디아
롯데백화점 청량리점 OGG_ECONOMY	<p>롯데백화점 청량리점</p> <p>위키피디아, 우리 모두의 백과사전</p> <p>롯데백화점 청량리점(百貨店 淸凉里店)은 롯데백화점에 속한 지점으로 롯데백화점의 네 번째 점포이다. 대한민국 서울특별시 동대문구 왕산로 200(전 동종)에 위치해 있으며 1994년 3월 18일에 개점하였으나 1996년 4월에 화재로 휴장하게 되었고, 같은 해 10월에 재장업을 실시하였다.</p>

● 개체명이 하위 조직을 의미하는데 해당 하위 조직이 별도의 문서로 존재하지 않는 경우 상위 조직 문서를 연결한다.

○ 예제

- (하위 조직) 트위터 코리아
- <트위터 코리아: OGG_ECONOMY>는 글로벌 운영 정책을 따르고 있다.
- (개체명) 트위터 코리아: OGG_ECONOMY → (위키피디아) 트위터

<p>개체명</p>	<p>위키피디아</p>
<p>트위터 코리아</p> <p>OGG_ECONOMY</p>	<div data-bbox="619 1243 1391 1377"> <div>트위터</div> <div>위키백과, 우리 모두의 백과사전.</div> <div>  트위터에 대해서는 트위터 소위키 문서를 참고하십시오. </div> <div>트위터(영어: Twitter)는 소셜 네트워크 서비스이자 마이크로블로그 서비스이다.</div> <div> <div>특징</div> <div>[보기]</div> </div> </div> <div data-bbox="619 1400 1160 1599"> <div>역사</div> <div>[편집]</div> <div> <div>이 문단은 아직 미완성입니다.</div> <div>여러분의 지식으로 알차게 문서를 완성해 갑시다.</div> </div> <div> <div>창업 초기</div> <div>[편집]</div> <p>트위터의 창업주는 잭 도시, 비즈 스톤, 에번 윌리엄스(CEO, 노아 글래스이다. 트위터는 미국 샌 프란시스코 지역의 팟캐스트 벤처 기업인 오데오(Odeo, Inc.)의 역변 윌리엄스와 노아 글래스가 고안한 연구 개발 프로젝트에서 출발하였다.^[1] 잭 도시는 SMS를 통해 소규모 그룹과 소통할 수 있는 방법을 고안해냈다. 최초의 프로젝트 코드네임은 twttr로 정해졌다. 당시 twitter.com이라는 도메인이 이미 사용중이었기 때문에 있는대, 6개월 후에 도메인을 구입했고, 서비스명을 트위터(Twitter)로 변경했다.^[2]</p> <p>트위터는 후버 존 레일즈를 사용하여 개발되었다.^[3]</p> <p>2006년 3월에 서비스를 시작하였다.</p> </div> </div> <div data-bbox="1168 1352 1391 1646"> <div>트위터</div> <div>  </div> <div> <div>분야</div> <div>URL</div> </div> <div> <div>영리여부</div> <div>영리</div> </div> <div> <div>사이트 종류</div> <div>소셜 네트워크 서비스</div> </div> <div> <div>최종 개발</div> <div>필요</div> </div> <div> <div>사용 언어</div> <div>한국어</div> </div> <div> <div>사용자</div> <div>약 3억 9천 5백만명 (2016년 기준)</div> </div> <div> <div>작성한 언어</div> <div>자바, 루비, 스칼라, 자바스크립트</div> </div> <div> <div>소유자</div> <div>잭 도시</div> </div> <div> <div>제작자</div> <div>잭 도시, 노아 글래스, 비즈 스톤, 에번 윌리엄스</div> </div> <div> <div>시작일</div> <div>2006년 3월 21일 (15년 전)</div> </div> <div> <div>현재의 상태</div> <div>> 30억회 / 1년 세계, 2021년 8월</div> </div> </div>

● 즉, ‘이마트 춘천점’, ‘신세계백화점 강남점’, ‘아메리칸 어패럴 코리아’, ‘구글 코리아’ 등과 같은 하위 조직에 대한 문서가 별도로 존재하면 해당 문서를 연결하고, 하위 조직 문서가 별도로 존재하지 않으면 예외적으로 상위 조직 ‘이마트’, ‘신세계백화점’, ‘아메리칸 어패럴’, ‘구글’ 문서를 연결한다.

● 이외의 경우에는 연결하지 않는다.

2.2.4. 개체명이 전체어를 가지는 경우

(1) 개체명이 부분어인데, 위키피디아에서 해당 개체명을 검색하였을 때 전체어 문서로 자동으로 넘어가는 경우, 연결하지 않고 '없음'에 해당하는 정보를 표시하는 것을 원칙으로 한다. 이에 해당하는 사례는 다음과 같다.

- 노래 제목 → 앨범명
- 가수 이름 → 그룹명
- 캐릭터 이름 → 캐릭터 목록

(2) 단, 시즌제 TV 프로그램이나 드라마의 경우에는 각각의 시즌이 별도의 문서로 존재하지 않고 대표 문서 안에서 각각의 시즌을 기술하고 있다면, 예외적으로 해당 시즌을 대표 문서에 연결한다.

- 예제
 - (TV 프로그램) 1박 2일 시즌4
 - <1박 2일 시즌4: AFA_VIDEO> 김선호 없었으면 어쩔 뻔!
 - (개체명) 1박 2일 시즌4: AFA_VIDEO → (위키피디아) 1박 2일

개체명	위키피디아
1박 2일 시즌4 AFA_VIDEO	

- 영화 시리즈물의 경우에는 연결하는 사례에 해당하지 않는다. 영화 시리즈는 각각의 시리즈가 독립적인 작품이다. 따라서 해당 영화 시리즈가 별도의 문서로 존재하면 연결하고, 별도의 문서로 존재하지 않으면 '없음'에 해당하는 정보를 표시한다.

2.2.5. 개체명이 동일지시적 표현을 가지는 경우

(1) 정식 명칭의 일부만 표현된 개체명

- 개체명 태깅 지침상 문맥을 고려했을 때 특정 개체명의 일부로 해당 개체를 가리킨다고 판단되는 경우, 복원된 형태에 준하여 문맥의 의미에 해당하는 세분류로 태깅하는 것으로 되어 있다. 그러나 위키피디아는 정식 명칭으로 문서를 제공하고 있으므로 정식 명칭의 일부만 표현된 제목의 문서는 존재하지 않는다. 따라서 온전한 형태를 제목으로 한 위키피디아 문서를 찾아 연결한다.

○ 예제

- (지하철역) 오목교
- 5호선 타고 오시다가 <오목교: LC_OTHERS>에서 내리시면 돼요.
- (개체명) 오목교: LC_OTHERS → (위키피디아) 오목교역

개체명	위키피디아
오목교 LC_OTHERS	<p>오목교역</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>오목교(목동운동장앞)역(Omokgyo (Mokdong Stadium) station, 梧木橋(木洞運動場앞)驛)은 서울특별시 양천구 오목로에 있는 수도권 전철 5호선의 지하철역이다. 목동운동장앞은 서울시 고시로 지정한 병기역명이다.^[1]</p>

○ 예제

- (단체) 수원 삼성
- <수원 삼성: OGG_SPORTS>이 인천 유나이티드를 3대 0으로 꺾었다.
- (개체명) 수원 삼성: OGG_SPORTS → (위키피디아) 수원 삼성 블루윙즈

개체명	위키피디아
수원 삼성 OGG_SPORTS	<p>수원 삼성 블루윙즈</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>수원 삼성은 여기로 연결됩니다. 과거 수원을 연고로 했던 프로농구단에 대해서는 서울 삼성 연다스 문서를 참조하십시오.</p> <p>수원 삼성 블루윙즈는 1995년 12월 15일에 창단한 대한민국의 K리그1 소속 프로축구단이다. 연고지는 경기도 수원시이며, 홈 구장은 '백색대리' 불리는 수원월드컵경기장이다.</p> <p>수원은 1995년 삼성전자가 수원시와 연고 협약을 맺고 K리그의 9번째 구단으로 출범했다. 수원은 1996 시즌부터 K리그에 참가해 정규리그 준우승과 FA컵 준우승을 기록했다. 1998년 정규리그에서 첫 우승에 성공했다. 1999년에는 전관왕을 달성하여 위용을 떨쳤다.</p> <p>2014년 4월 1일부터 구단 모기업이 삼성전자에서 같은 삼성그룹의 계열사인 제일기획으로 이관됐으며, 삼성전자는 모기업에서 스폰서로 전환하여 구단의 운영과 지원이 이원화됐다. 이는 스포츠 전문 기업이 아닌 제조회사의 삼성전자에 속해 그동안 효율적이지 못한 구단 운영의 한계에서 벗어나 그동안 수원 마케팅을 담당한 삼성그룹 계열사 제일기획에서 운영함으로써 그동안 성적 위주의 운영에서 벗어나 스포츠사업으로써 이익을 추구하는 방향으로 바뀌어 구단의 운영과 지원을 효율적으로 할 수 있는 이원화 방식을 채택한 것^[1]과 삼성스포츠단 재편으로 인해 삼성스포츠단 산하 스포츠팀을 전부 제일기획에서 관리할 예정에 이뤄진 조치이다. 이와 함께 법인명도 삼성전자축구단에서 수원 삼성 블루윙즈 축구단으로 바뀌었다.</p> 

(2) 줄임말

- 개체명이 줄임말인 경우에도 복원된 형태에 준하여 문맥의 의미에 해당하는 세분류로 태깅한다. 그러나 위키피디아는 앞서 설명한 바와 같이 정식 명칭으로 문서를 제공하고 있으므로 줄임말을 검색하면 정식 명칭 문서로 넘어간다. 따라서 정식 명칭의 일부만 표현된 개체명과 마찬가지로 줄임말인 개체명도 위키피디아의 정식 명칭 문서를 연결한다.

○ 예제

- (기관) 여가부
- <여가부: OGG_POLITICS>는 지난해 11월부터 카카오톡으로 성범죄자 신상정보를 제공하기 시작했다.
- (개체명) 여가부: OGG_POLITICS → (위키피디아) 대한민국 여성가족부

개체명	위키피디아
<p>여가부 OGG_POLITICS</p>	

○ 예제

- (게임) 워3
- <워3: AFW_OTHER_PRODUCTS> 황제 장재호가 한-중 매치 승리를 견인했다.
- (개체명) 워3: AFW_OTHER_PRODUCTS → (위키피디아) 워크래프트 III: 레인 오브 카오스

개체명	위키피디아
<p>워3 AFW_ OTHER_PRODUCTS</p>	


(3) 일상적으로 부르는 말

- 개체명 태깅 지침상 정식 명칭이 아닌 일상적으로 부르는 명칭 또한 문맥의 의미에 해당하는 세분류로 태깅하는 것으로 되어 있다. 그러나 위키피디아는 정식 명칭으로 문서를 제공하고 있어 일상적으로 부르는 명칭을 검색할 경우, 정식 명칭으로 기술된 문서로 넘어가게 되어 있다. 위키피디아에 일상적으로 부르는 명칭의 문서가 따로 존재하지 않으므로 정식

명칭 문서를 연결한다.

○ 예제

- (기업) 콩다방
- 바닐라라떼는 역시 <콩다방: OGG_FOOD>이지
- (개체명) <콩다방: OGG_FOOD> → (위키피디아) 커피빈 & 티리프

개체명	위키피디아
<p>콩다방 OGG_FOOD</p>	<p>커피빈 & 티리프</p> <p>위키백과, 우리 모두의 백과사전</p> <p>(콩다방에서 보기)</p> <p>커피빈 & 티리프(영어: The Coffee Bean & Tea Leaf, '커피 콩과 차 잎'이라는 뜻) 또는 커피빈은 미국 로스앤젤레스에 본사가 있는 인터내셔널 커피 & 차 유한책임회사(International Coffee & Tea, LLC)에서 운영되고 있는 세계에서 가장 오래된 커피 및 홍차 전문점 프랜차이즈 중의 하나이다. 최초의 매장은 1963년에 로스앤젤레스의 브랜트우드에서 허버트 & 헤이먼과 도나 헤이먼이 개점하였다. 현재 미국 남서부와 하와이주까지 아우르는 점포망을 갖고 있다. 동남아시아에는 1996년에 출판권을 구입한 싱가포르 출신의 사업가 벡터 세운이 '커피빈스'(Coffee Beans)라는 이름의 매장을 여러 곳에 개점하였다. 전 세계적으로는 현재 14여 개국, 400여 곳이 넘는 직영 및 프랜차이즈 매장을 갖고 있다.</p> <p>커피빈은 커피빈 여러 여러 종류의 홍차 음료를 취급한다는 점에서 스타벅스 같은 다른 커피 전문점과는 다르다. 또, 유명한 오리지널 아이스 블렌디드(Original Ice Blended)후 블렌드와 자이라에도 취급한다.</p> <p>2009년에는 매출 1억 2천만 달러(USD)를 기록하였다. 점포 중 144여 곳에서는 와이파이가 접속을 지원한다. 미국 외에서는 주로 프랜차이즈 매장이며, 미국 내에는 주로 회사 소유의 매장이 위치한다. 현재 진출 국가는 싱가포르, 이스라엘, 대한민국, 아랍 에미리트, 브루나이, 오스트레일리아, 인도네시아, 말레이시아, 필리핀, 쿠웨이트, 이집트이다.[1]</p> <p>2019년 9월 필리핀의 식품기업 줄리베(Jollibee Foods)가 커피빈의 브랜드를 보유한 International Coffee & Tea, LLC를 인수했다.</p> <div> <div> <p>커피빈 & 티리프 The Coffee Bean & Tea Leaf</p>  </div> <div> <p>산업 분야 음식점</p> <p>창립 1963년, 로스앤젤레스 브랜트우드 (로스앤젤레스)</p> <p>국가 미국</p> <p>본사 소재지 필리핀마닐라 로스앤젤레스</p> <p>핵심 인물 COO: 커나 세운 COO: 알빈 엘리엇스</p> <p>제품 커피 빈 앤 차 리츠 등</p> </div> </div>

○ 예제

- (국가) 북한
- <북한: LCP_COUNTRY> 주민들이 즐겨 쓰는 스마트폰 앱은?
- (개체명) <북한: LCP_COUNTRY> → (위키피디아) 조선민주주의인민공화국

개체명	위키피디아
<p>북한 LCP_COUNTRY</p>	<p>조선민주주의인민공화국</p> <p>위키백과, 우리 모두의 백과사전</p> <p>(북한에서 보기)</p> <p> 북한은 여기로 연결됩니다. 중국 초대입국 시대 10국 중 하나에 대해서는 북한 (입국) 문서를 참조하십시오.</p> <p>조선민주주의인민공화국(朝鮮民主主義人民共和國, 영어: Democratic People's Republic of Korea, DPRK) 또는 줄여서 북조선, 조선은 동아시아의 한반도의 북반부를 통치하고 있는 국가이다.[1] UN이 추정하는 조선민주주의인민공화국의 인구는 2019년 기준으로 약 2,550만 명이다.[2] 조선민주주의인민공화국의 공용어는 조선어이며, 평양말을 중심으로 아랍어를 내세운 문화어를 표준어로 삼는다.</p> <p>수도는 평양직할시이며, 어번에 조선민주주의인민공화국의 공적자치체로는 남포특별시, 개성특별시, 라선특별시 등이 있다.</p> <p>조선민주주의인민공화국의 통치 형태는 형식상 정당이 여러 존재하지만, 대부분의 정당이 조선로동당의 정책과 유사한 관례로 사실상 일당제 국가라는 평가를 받는다. 초대 최고지도자 김일성은 6.25 전쟁, 8월 종파 사건 등을 거쳐 정적을 제거하여 수령 중심 정치 체제를 완성하였다. 그의 자손인 김정일, 김정은은 차례로 집권을 하여 선대의 수령주의를 이어받은 정치를 펼쳤다. 이 체제는 '주체사상'이라는, 김일성 일가와 그들의 당파인 조선로동당의 통치를 정당화하는 이념으로 대표된다.</p> <div> <div> <p>조선민주주의인민공화국</p>  </div> <div>  </div> </div> <div> <p>국기</p> <p>국장</p> <p>모어</p> <p>차려문자, 공산대국어[3]</p> <p>국가</p> <p>애국가</p> </div>

- 정식 명칭의 일부만 표현된 말, 줄임말, 일상적으로 부르는 말 외에도 하나의 대상을 지시하는 언어 표현이 다수 존재할 수 있는데, 개체명을 태깅할 때는 이렇게 동일지시적인 언어 표현을 정식 명칭으로 수정하지 않고 해당 언어 표현 그대로에 개체명 태그를 부착한다. 그러나 위키피디아는 하나의 대상을 지칭하는 표현이 다수 존재할 때 언어 표현별로 제목을 달리하여 문서를 따로 제공하는 것이 아니라, 최신의 정식 명칭을

제목으로 하는 문서 하나를 제공한다. 그러므로 해당 대상에 대해 위키피디아에서 제공하는 정식 명칭의 문서를 연결해야 한다. 정리하면 동일지시적인 개체명들은 다음과 같이 하나의 위키피디아 문서와 연결될 수 있다.

○ 예제

- 방탄소년단의 리더를 지칭하는 말: 김남준, 랩몬스터, 랩몬, RM, 알엠
 - <김남준: PS_NAME>의 매력은 끝이 어디일까요?
 - 제가 방탄소년단 <랩몬스터: PS_NAME>를 좋아하는데요.
 - BTS 영상을 열심히 보던 안현모는 <랩몬: PS_NAME>의 영어에 흠족해했다.
 - 방탄소년단 리더 <RM: PS_NAME>의 사진 한 장이 아이들의 마음을 설레게 하고 있다.
 - <알엠: PS_NAME>은 학창 시절부터 미국 드라마 등을 통해 자연스러운 감을 익혔다고 전했다.
- (개체명) 김남준, 랩몬스터, 랩몬, RM, 알엠: PS_NAME → (위키피디아) RM (래퍼)

개체명	위키피디아
김남준 랩몬스터 랩몬 RM 알엠 PS_NAME	<div> <div>RM (래퍼)</div> <div>한글 40개 언어 ▾</div> <div> <p>위키백과, 우리 모두의 백과사전.</p> <p>RM(본명: 김남준, 金南俊^{[*]}, 1994년 9월 12일 ~)은 대한민국의 래퍼이자 작사가, 작곡가, 프로듀서이다. 대한민국의 보이 그룹 방탄소년단의 리더, 메인래퍼를 맡고 있다.^{[*]} 2013년부터 약 5년간 랩몬스터(영어: Rap Monster)라는 예명을 사용했으나, 2017년 11월 13일부터 'RM'으로 예명을 바꿔 활동하고 있다. 그리고 2015년, 첫 솔로 믹스테이프<Rap Monster>을 발표하였다. 100여 곡이 한국음악저작권협회에 등록되어 있으며,^{[*]} 방탄소년단과 라인프렌즈가 함께 만든 캐릭터인 BT21 중에서 RM은 코야(KOYA)라는 캐릭터를 직접 디자인했다.데뷔 전에는 '런치란다'라는 예명을 이용해 언더래퍼로 활동한 바 있다.</p> </div>  </div>

2.3. 개체명의 대분류에 따른 세부 원칙

2.3.1. PERSON

(1) 인명 및 인물의 별칭

- 개체명이 유명 인사인 것이 맥락을 통해 명확하게 드러나는 경우에만 위키피디아 문서를 연결하고, 맥락을 통해 개체명이 가리키는 대상이 누구인지를 명확하게 판별할 수 없는 경우에는 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

○ 예제

- (인명) 수지
- 국민 첫사랑 '건축학개론'의 <수지: PS_NAME>
- (개체명) 수지: PS_NAME → (위키피디아) 수지 (1994년)

개체명	위키피디아
수지 PS_NAME	<p>수지 (1994년) 총 39개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p> 백수지는 여기로 연결됩니다. 다른 뜻에 대해서는 백수지 (동음이의) 문서를 참조하십시오.</p> <p>수지(본명: 배수지, 1994년 10월 10일 ~)는 대한민국의 前 가수이자 배우이다.</p> <p>2009년 Mnet 슈퍼스타K 1 광주 예선에 참여했다가 JYP 관계자에게 캐스팅 되어 JYP 엔터테인먼트의 연습생이 되었다. 이후 1년 동안의 연습생 시절을 거쳐 2010년 과거 4인조 걸 그룹 미쓰에이의 멤버로 데뷔했다. 2011년 드라마 《드림하이》로 연기자로 입문하였고, 이후 주요 출연작으로는 드라마 《빅》 (2012), 《구가의 서》 (2013), 《함부로 애틋하게》 (2016), 《당신이 잠든 사이에》 (2017), 《배가본드》 (2019)와 영화 《건축학개론》 (2012), 《도리화가》 (2015), 《백두산》 (2019)이 있다.</p> 


○ 예제

- (인명) 수지
- <수지: PS_NAME> 너무 예쁘다.
- (개체명) 수지: PS_NAME / (위키피디아) NONE (연결 X)
- 개체명이 '성 + 직위/직책' 구성인 경우에도 맥락을 통해 개체명이 지시하는 대상을 명확하게 판별할 수 있는 경우에는 해당 성명(성+이름)을 제목으로 하는 위키피디아 문서를 연결한다. 그러나 맥락을 통해 개체명이 가리키는 대상이 누구인지 명확하게 판별할 수 없는 경우에는 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

○ 예제

- (인명) 김 추기경
- <김수환: PS_NAME> 추기경 선종 10주기에 맞춰 추모 LP앨범이 발매된다. 이 한정판 성가집에는 <김: PS_NAME> 추기경님께 보내는 이해인 수녀님, 법정스님 등의 글을 비롯하여 박노해 시인의 추모시 '거룩한 바보'도 수록된다.
- (개체명) 김: PS_NAME → (위키피디아) 김수환

개체명	위키피디아
김	

PS_NAME	<div> <div>김수환</div> <div>위키백과, 우리 모두의 백과사전.</div> <div> <div>같은 이름을 가진 동명의 아나운서에 대해서는 김수환 (아나운서) 문서를 참조하십시오.</div> <div>같은 이름을 가진 동명의 야구 선수에 대해서는 김수환 (야구 선수) 문서를 참조하십시오.</div> </div> <div> <div>김수환(金壽煥, 1922년 7월 2일 (음력 5월 8일)^[1] ~ 2009년 2월 16일)은 대한민국의 천주교 성직자이자 사회운동가이다. 한국인 최초로 가톨릭 주교령에 서임되었다. 본관은 공산. 세례명은 스테파노이며 아호는 뽕기이다.</div> <div> <div>박정희, 전두환, 노태우 군사 정권 당시 군사 독재정권 퇴진 운동과, 시민 활동을 하였다. 문인정부 출범 이후부터는 사회 운동과 방송 활동, 복지 사업, 언론 활동, 강연 활동 등을 하였다. 한국 가톨릭계를 대표하는 인물로서 수십 년간 군부 정권의 독재에 저항하며 한국 민주주의의 발전을 위해 헌신한 인물 가운데 한 사람이었다.^[2] 인권의 수호자로서 인간의 존엄성에 대한 신념과 공동선의 추구를 바탕으로 교회가 세상에서 빛과 소금의 역할을 감당해야 한다는 신념에 따라 신앙을 실천한 인물이었다.^[3] 로마교구의 '산 펠리체 다 칸탈리체 천토렐레 성당' 영예주임사제직의 사제급 추기경이었다.</div> <div> <div>스테파노 김수환 사제급 추기경</div>  </div> </div> </div> </div>
---------	--

○ 예제

- (인명) 이 회장
- 이번 사태에 대한 <이: PS_NAME> 회장의 결단이 기대된다.
- (개체명) 이: PS_NAME / (위키피디아) NONE (연결 X)

- 개체명이 '성 + 씨' 구성인 경우에도 맥락을 통해 개체명이 지시하는 대상을 명확하게 판별할 수 있는 경우에는 해당 성명(성+이름)을 제목으로 하는 위키피디아 문서를 연결한다. 맥락을 통해 '성 + 씨' 구성이 가리키는 대상이 누구인지 명확하게 판별할 수 없는 경우에는 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다. 위키피디아에 '성씨' 문서가 별도로 존재하는 경우가 있는데, 이 '성씨' 문서를 연결해서는 안 된다.

○ 예제

- (인명) 박 씨
- 김(범수), 나(얼), 박(효신), 이(수) 중에서 <박: PS_NAME> 씨가 탑.
- (개체명) 박: PS_NAME → (위키피디아) 박효신

개체명	위키피디아
<div>박</div> <div>PS_NAME</div>	<div> <div>박효신</div> <div>위키백과, 우리 모두의 백과사전.</div> <div> <div>박효신(朴孝信, 1981년 9월 1일~)은 대한민국의 남성 솔로가수이다.</div> <div> <div>목록 [보이기]</div> </div> <div> <div>약력 [편집]</div> </div> <div> <div>박효신은 고등학교 1학년 재학 중에 '무전 청소년 가요제'에서 이기찬의 〈Please〉로 대상을 차지하고 〈YMCA 청소년 가요제〉, 〈재물포 가요제〉 등의 대회에서 수상하면서 가수의 꿈을 키웠다. 1999년 11월 4일 SBS 라디오 〈영스트리트〉 공개방송에서 데뷔하였고, 첫 음반 〈해줄 수 없는 일〉은 같은 해 1999년 12월에 발매되었다.</div> <div> <div>박효신</div>  </div> </div> </div> </div>

○ 예제

- (인명) 임 씨

- <임: PS_NAME> 씨가 정원을 잘 꾸며 놓은 듯.
 - (개체명) 임: PS_NAME → (위키피디아) NONE (연결 X)
- 동일 음절 중복으로 두 인명이 이어진 경우, 둘 중 한 명의 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (인명) 이명박근혜
 - <이명박근혜: PS_NAME>라는 말은 박근혜 본인이 처음 사용했다.
 - (개체명) 이명박근혜: PS_NAME / (위키피디아) NONE (연결 X)
- @sultan_of_the_disco와 같은 SNS 계정명의 경우, 실제 인물과 계정명을 연결하지 않는다. 유명인의 SNS는 운영자가 따로 있는 경우가 있을 뿐만 아니라 SNS는 인물 자체가 아닌, 해당 인물이 소유하고 있는 대상이므로 동일지시적이라 볼 수 없다. 따라서 이러한 SNS 계정명은 위키피디아에 동명의 인물 문서가 존재한다고 하더라도 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (계정명) @sultan_of_the_disco
 - <@sultan_of_the_disco: PS_NAME> 보세요.
 - (개체명) @sultan_of_the_disco: PS_NAME / (위키피디아) NONE (연결 X)
- 어떤 인물이 음악 그룹에 속해 있는데, 해당 인물을 위키피디아에 검색하였을 때 인물에 대한 문서가 별도로 존재하지 않고 음악 그룹으로 문서가 넘어가는 경우에는 개체명 ‘인물’에 위키피디아 ‘음악 그룹’ 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (인명) 제이어스
 - <제이어스: PS_NAME>한테 빠졌어요.
 - (개체명) 제이어스: PS_NAME / (위키피디아) NONE (연결 X)

(2) 작품 속 등장인물의 이름

- 영화나 드라마 속 등장인물의 이름을 위키피디아 문서와 연결할 때에는 해당 캐릭터에 관한 위키피디아 문서를 연결한다. 해당 작품 속 캐릭터를 연기한 배우의 위키피디아 문서를 연결하지 않는다. 해당 캐릭터명이 제목인 위키피디아 문서가 존재하지 않는다면 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (가공의 인물) 아이언맨
 - <아이언맨: PS_CHARACTER> 죽지 마.
 - (개체명) 아이언맨: PS_CHARACTER → (위키피디아) 아이언맨

개체명	위키피디아
아이언맨 PS_CHARACTER	<p>아이언맨</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>본 문서의 대역사는 아이언맨 (동음이의) 문서를 참고하십시오.</p> <p>아이언맨(영어: Iron Man)은 미국 만화인 마블 코믹스에 등장하는 슈퍼 히어로이다. 스탠 리, 레리 리 에버가 줄거리를 담당하고 론 페크와 잭 키리가 그들을 맡았다. 아이언맨은 1963년 3월 서스펜스 이야기 39호에 처음으로 등장한다.</p> <p>부유한 사상계의 거물이자 배달물이, 기발한 공학자, 천재인 엔지니어 에드워드 스타크(영어: Anthony Edward Stark)는 대량살상무기를 실험하던 중 남자를 당해 가슴에 심박전 부상을 입었다. 그는 자신의 복수를 실행할 수 있는 전자로 수트를 입고 활물에 성공한다. 이후 스타크는 그의 회사인 스타크 인더스트리에서 디자인한 무기과 다른 기술적 장치들을 수트에 장착한다. 그는 아이언맨으로서 세계를 보호하기 위해 수트와 그 이후 개량 버전들을 사용한다. 그러나 그는 처음에 진짜 자신의 정체성을 숨겼다. 초기에 스타크 리는 아이언맨을 냉전과 관련시켜 만들었고, 미국의 기술과 산업의 역할을 공산주의와의 싸움에 이용하는 것을 표현했다.^[1] 아이언맨의 이후 이미지는 재탄생되어 냉전의 주제를 벗어나 테러리즘과 회사범죄를 다루기 시작했다.^[2]</p> <p>초기에 아이언맨은 스타크 리에게 냉전의 주제, 특히 공산주의와의 싸움에서 미국의 기술과 산업의 역할을 탐구하는 매개체였다. 그러나 이후 아이언맨은 냉전의 묘미에서 당시의 현대사안으로 옮겨갔다.^[3]</p> <p>캐릭터 출판 기간 중 대부분의 시간, 아이언맨은 어벤져스의 창립 회원이었고, 그의 만화책 시리즈에서 여러 생을 살기도 했다. 호브트 다운나 주니어가 아들 시네마틱 유니버스에서 이 캐릭터를 맡았다. 2008년 아이언맨을 시작으로, 그는 인크레더블 필름에서 카메오를 맡았으며, 이후 2편의 아이언맨 영화인 아이언맨 2와 아이언맨 3에서 주인공으로 등장했다. 이후 어벤져스: 어벤져스: 에이지 오브 울트론, 캡틴 아메리카: 시빌 워, 스파이더맨: 홈커밍, 어벤져스: 인피니티 워, 어벤져스: 엔드게임에서도 이 역할을 맡았다. 닥터 스트레인지인 윌리엄스 오브 매도니스에서는 폴 크루즈가 역할을 맡는다.</p> <p>아이언맨은 R과 2011년 "100대 만화 영웅들"에서 12위를 차지했으며, 2012년 "어벤져스 50명" 중 3위를 차지했다.^[4]</p> 

○ 예제

- (가공의 인물) 이익준
- ‘슬기생2’의 슈퍼 인싸 <이익준: PS_CHARACTER>!
- (개체명) 이익준: PS_CHARACTER / (위키피디아) NONE (연결 X)
- tvN 드라마 ‘슬기로운 의사 생활 시즌2’의 등장인물인 ‘이익준’에 대한 문서는 위키피디아에 존재하지 않는다. ‘이익준’ 역의 ‘조정석’ 문서는 존재하는데, 캐릭터 ‘이익준’을 배우 ‘조정석’이 연기한다고 하여 ‘조정석’ 문서를 ‘이익준’에 연결해서는 안 된다. ‘이익준’과 ‘조정석’은 각각 가상의 인물과 실존 인물, 대상 자체가 다르므로 ‘이익준’에 위키피디아 문서 ‘조정석’을 연결하지 않는다.
- 국립국어원의 <2020년 개체명 분석 말뭉치 구축 지침>에서는 소설이나 영화, 드라마 등의 작품에 등장하는 실존 인물은 구문 맥락상 ‘작품 속의 인물’을 지칭하는 것이 명확하다면 PS_CHARACTER로 태깅하도록 하고 있다. PS_CHARACTER로 태깅된 실존 인물의 경우, 작품 속 인물에 대한 문서가 위키피디아에 별도로 존재하지 않는다면 실존 인물 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (작품 속 인물) 마크 저커버그
- 마크 저커버그는 영화 ‘소셜 네트워크’ 속 <저커버그: PS_CHARACTER>가 여자 친구에게 차여서 facemash.com을 차리는 부분은 사실이 아니라고 해명했다.
- (개체명) 저커버그: PS_CHARACTER / (위키피디아) NONE (연결 X)

2.3.2. ARTIFACTS

(1) 동명의 작품명

- 동명의 소설, 그림, 영화, 음악 등이 존재하는 경우, 즉 서로 다른 예술 분야에 동명의 작품이 존재하는 경우에는 개체명 태깅 범주와 동일한 예술 분야의 문서를 연결한다.

○ 예제

- (작품명) 어린 왕자
- 읽을 때마다 새로운 책, <어린 왕자: AFA_DOCUMENT>
- (개체명) 어린 왕자: AFA_DOCUMENT → (위키피디아) 어린 왕자

개체명	위키피디아
<p>어린 왕자 AFA_DOCUMENT</p>	<p>어린 왕자 총 108개 언어</p> <p>위키백과, 우리 모두의 백과사전. (어린왕자에서 넘어옴)</p> <p>다른 뜻에 대해서는 어린 왕자 (동음이의) 문서를 참고하십시오.</p> <p>《어린 왕자》(프랑스어: Le Petit Prince)는 프랑스의 비행사이자 작가인 앙투안 드 생텍쥐페리가 1943년 발표한 소설이다.</p> <p>1943년에 미국에서 처음 출판되었고, 그 해 비시 프랑스 치하의 프랑스에서 비밀리에 출판되었다. 프랑스가 해방된 이후 1947년 가리마르사(社)가 작가 자신의 이상하고 아름다운 삽화를 넣어 프랑스에서 새로 출판하였다.^[1]</p> <p>현재까지 300여 개 국어로 번역되었고^[2], 한국어판 중에는 저자의 삽화가 삽입되어 있는 번역본이 있다. 영화와 애니메이션으로도 제작되었다.</p> 

○ 예제

- (작품명) 어린 왕자
- 애니메이션 <어린 왕자: AFA_VIDEO>에서 여우 역은 프랑스 배우 뱅상 카셀이 더빙했다.
- (개체명) 어린 왕자: AFA_VIDEO → (위키피디아) 어린 왕자 (2015년 영화)

개체명	위키피디아
<p>어린 왕자 AFA_VIDEO</p>	<p>어린 왕자 (2015년 영화)</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>《어린 왕자》(프랑스어: Le Petit Prince)는 2015년 공개된 프랑스의 애니메이션 영화이다.</p>

- 어떤 작품이 드라마, 영화로 재구성되었거나 동명의 서로 다른 작품이 존재하는데, 대상 문서의 맥락을 통해서는 개체명이 가리키는 작품이 무엇인지 명확하게 판별할 수 없는 경우, 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (작품명) 더블유
- <더블유: AFA_VIDEO> 재밌더라.
- (개체명) 더블유: AFA_VIDEO / (위키피디아) NONE (연결 X)

- ‘더블유’라는 제목의 미국 영화가 있고, 동명의 한국 드라마가 있다. ‘더블유 재밌더

라.’라는 문장에서 ‘더블유’가 영화인지 드라마인지 특정할 수 없으므로, 위키피디아에 ‘더블유 (영화)’ 문서와 ‘더블유 (드라마)’ 문서가 각각 존재하지만 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

(2) 시리즈로 제작된 영화

- 영화 ‘스타워즈’, ‘어벤져스’ 등과 같이 시리즈로 제작된 영화의 경우, 각각의 시리즈는 독립적인 작품이라고 볼 수 있다. 따라서 영화의 특정 시리즈가 위키피디아에 별도의 문서로 존재하는 경우에는 해당 개체명과 위키피디아 문서를 연결한다. 특정 시리즈가 위키피디아에 별도의 문서로 존재하지 않는 경우에는 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (작품명) 스타워즈: 깨어난 포스
- <스타워즈: 깨어난 포스: AFA_VIDEO> 완전 기대.
- (개체명) 스타워즈: 깨어난 포스: AFA_VIDEO → (위키피디아) 스타워즈: 깨어난 포스

개체명	위키피디아
스타워즈: 깨어난 포스 AFA_VIDEO	<p>스타워즈: 깨어난 포스</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>《스타워즈: 깨어난 포스》(Star Wars: The Force Awakens)는 2015년에 공개된 미국의 스페이스 오페라 영화이다. J.J. 에이브람스가 감독과 공동 각본, 공동 제작을 맡았다. 《스타워즈》 영화 시리즈의 7번째 작품으로 데이비지 리틀리, 존 보에가, 애덤 드라이버, 오스카 아이작, 제리슨 모도, 캐리 피셔, 루피타 뇽오, 앙디 서키스, 도널 글리슨, 맨서니 대니얼스, 피타 데이비스, 맥크 제일, 맥스 폰 위도브가 출연하였다. 루카스필름과 에이브람스의 제드 로브트 프로덕션스가 제작하였고 월트 디즈니 스튜디오스 모션 픽처스가 전세계 배급을 담당하였다. 《제다이의 귀환》 이후 약 30년 뒤를 시간적 배경으로 하여 레이와 핀, 반란동맹군의 군인들이 이끄는 저항군의 로 디에온이 루크 스카이워커를 찾고, 카일로 연과 온하 제국을 계승한 퍼스트 오더와 싸우는 과정을 그리고 있다.</p> <p>2012년 10월 스타워즈의 판권이 루카스필름에서 월트 디즈니 컴퍼니에 넘어간 뒤 새로운 에피소드를 다룬 3부작이 계획되었고 이 영화를 시작으로 두 편의 후속작이 더 나올 예정이다. 촬영은 2014년 4월 아랍 에미리트 아부다비 도주국과 아예슬란드에서 시작되었으며 주요 촬영은 아일랜드와 잉글랜드의 파인우드 스튜디오에서 이루어졌다.</p> <p>《깨어난 포스》는 미국 개봉일 4월 17일 전인 2015년 12월 14일 로스앤젤레스에서 처음 공개되었다. 2016년 1월 23일 기준으로 전세계에서 \$1,888,548,249의 수익을 기록하며 2015년 영화 매출 순위 1위와 역대 영화 매출 순위 3위를 기록하였다. 두 편의 후속편 《에피소드 8》과 《에피소드 9》가 각각 2017년 12월 15일과 2019년에 공개될 예정이었으나 한국에서는 2020년 1월에 공개되었다.</p> 

(3) 시즌제 TV 프로그램이나 드라마

- 미국 드라마 ‘프렌즈’나 예능 프로그램 ‘신서유기’, ‘1박 2일’ 등과 같이 시즌제로 제작되는 TV프로그램이나 드라마의 경우, 각각의 시즌이 별도의 문서로 존재하지 않고 하나의 문서 안에 관련 정보로서 포함되어 있는 경우, 예외적으로 해당 시리즈를 대표 문서에 연결한다.

○ 예제

- (작품명) 신서유기
- <신서유기4: AFA_VIDEO>가 제일 재밌었어.
- (개체명) 신서유기4 : AFA_VIDEO → (위키피디아) 신서유기

개체명	위키피디아
-----	-------



(4) 곡명과 음반명

- 노래 제목을 위키피디아에 검색하였을 때, 해당 노래를 주제로 하는 위키피디아 문서가 존재하지 않고 해당 노래가 수록되어 있는 음반으로 문서가 넘어가는 경우가 있다. 곡명은 AFA_MUSIC이나 음반명은 AFW_OTHER_PRODUCTS로 개체명 태그가 다르므로 이러한 경우에는 개체명 ‘노래 제목’에 위키피디아 ‘음반’ 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (곡명) Decalcomanie
- 마마무 <Decalcomanie: AFA_MUSIC> 신나!
- (개체명) Decalcomanie: AFA_MUSIC / (위키피디아) NONE (연결 X)

(5) 리메이크 곡

- 위키피디아에 리메이크 곡에 대한 문서가 존재하지 않고 원곡 문서만 있는 경우, 개체명 리메이크 곡에 위키피디아 원곡 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
- 이는 해당 리메이크 곡이 정식 음원, 음반으로 발매되었을 경우에 한하며, 방송 프로그램 등에서 단순히 리메이크하여 부른 경우에는 해당되지 않는다.

○ 예제

- (곡명) 뽀에로는 우릴 보고 웃지
- 아이유의 <뽀에로는 우릴 보고 웃지: AFA_MUSIC> 덕분에 알게 된 가수 김완선.
- (개체명) 뽀에로는 우릴 보고 웃지: AFA_MUSIC / (위키피디아) NONE (연결 X)
- 단순히 방송 프로그램 등에서 리메이크하여 부른 경우에는 원곡 문서를 연결한다.

○ 예제

- (곡명) Chandelier
- 박정현이 비긴어게인에서 <Chandelier: AFA_MUSIC> 부른 거 너무 좋아.
- (개체명) Chandelier: AFA_MUSIC → (위키피디아) Chandelier

개체명	위키피디아
Chandelier	

AFA_MUSIC	<div>Chandelier</div> <div>위키백과, 우리 모두의 백과사전.</div> <div>'Chandelier'는 오스트레일리아의 가수 시메의 노래로, 2014년 발매된 여섯번째 정규 음반 <i>1000 Forms of Fear</i>의 수록곡이다.</div> <div>외부 링크 [편집]</div> <div> <ul style="list-style-type: none">비디오클립 · 음원본 </div> <div> <div>"Chandelier"</div> <div>시메의 싱글</div> <div>1000 Forms of Fear의 싱글</div> <div>발매일</div> <div>2014년 3월 17일</div> <div>장르</div> <div>CD 싱글, 디지털 다운로드</div> </div>
-----------	---

(6) 지명 + 건물/기관명

- 건물/기관명 앞에 해당 건물/기관이 위치하고 있는 지역의 이름이 함께 나타나 있는 경우 다음과 같이 연결한다.

- 지명과 건물/기관명이 각각 따로 태깅되어 있는 경우에는 각각의 개체명에 위키피디아 문서를 연결한다.

■ 예제

- (건물) 한남동 디뮤지엄
- <한남동: LCP_COUNTY> <디뮤지엄: AF_BUILDING>에서 찍은 사진
- (개체명) 한남동: LCP_COUNTY → (위키피디아) 한남동
- (개체명) 디뮤지엄: AF_BUILDING → (위키피디아) 디뮤지엄

개체명	위키피디아
<p>한남동</p> <p>LCP_COUNTY</p>	<div>한남동</div> <div>위키백과, 우리 모두의 백과사전.</div> <div>한남동(漢南洞)은 서울특별시 용산구에 있는 동이다.</div> <div>특징 [보이기]</div> <div>개요 [편집]</div> <div>한남동의 위치는 동쪽으로는 성동구와 접하고, 서쪽으로는 이태원동과 보광동을 접하여 서북쪽으로는 남산을 주산으로하는 매봉산과 이어지고 있다. 그리고, 남쪽은 한강을 경계로 한 아름다운 강경을 바라볼 수 있고, 현재의 한남대로는 한남동을 양분시키고 있고, 이 길은 한남대교와 연결되어 남쪽의 고속도로로 달릴 수 있다. 한남동 726 번지 한남초등학교와 재봉대 군악대 주둔지를 예전에는 '농터골'이라고 했는데 이곳은 조선시대 때 농터로 여러 차례농은 자리라 해서 그렇게 불렀다고 한다. 또 부근에는 남산 밑에 있던 두 장사가 큰 배워들을 둘러다가 높은 '매송다리'라는 다리가 있었는데, 일제 강점기 때 조계사에서 가져가 벽석을 만드는 작업에 끌어갔다고 한다.^[1]</div> <div> <div>서울특별시 용산구</div> <div>한남동</div> <div>Hannam-dong</div>  </div>
개체명	위키피디아
<p>디뮤지엄</p> <p>AF_BUILDING</p>	<div>디뮤지엄</div> <div>위키백과, 우리 모두의 백과사전.</div> <div>좌표: 북위 37° 32′ 13″ 동경 127° 00′ 41″</div> <div>디뮤지엄 (D Museum)은 대한민국 서울에 위치한 미술관 및 전시관이다. 대림미술관의 부속관이자 대림문화재단에서 운영하는 곳 중 하나로, 2015년 12월에 한남동 독서당으로 개관했으나,^[1] 2021년 상수동 서울숲 인근으로 이전했다. 주조는 성동구 용입리로 83-21이다.</div>

- 개체명 태깅 지침상 건물/기관과 그것이 위치하고 있는 장소는 각각 따로 태깅하는 것으로 되어 있다. 이에 따라, 지명과 건물/기관명 전체가 한 단위로 태깅되어 있는 경우에는 위키피디아 문서를 연결하지 않고 'ERROR'로 처리한다.

■ 예제

- (건물) 여의도IFC몰
- <여의도IFC몰: AF_BUILDING> 3층에 CGV 있어.

● (개체명) 여의도IFC몰: AF_BUILDING / ERROR

- 단, '갤러리아백화점 광고점'과 같이 지점명이 함께 쓰인 경우에는 앞에서 기술한 '2.2.3. 개체명이 상위어-하위어 관계를 갖는 경우'에 의거하여 연결한다.
- '광주지방법원'과 같이 정식 명칭에 지명이 포함된 경우에는 개체명 태깅 지침대로 전체를 한 단위로 태깅하고, 위키피디아의 정식 명칭 문서인 '광주지방법원'을 연결한다.

(7) 기관(OGG)이 아닌 건물(AF)로 태깅되어 있는 기업의 하위 지점/지부

- 기업의 하위 지점이나 지부가 기관(OGG)의 의미를 가지는데, 해당 하위 지점/지부가 별도의 문서로 존재하지 않고 상위 기업 문서만 존재하는 경우에는 상위 기업 문서를 연결하나, 건물(AF)의 의미를 가질 때에는 상위 기업 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.
- AF로 태깅된 하위 지점/지부는 지리적 정보가 핵심적인 정보일 때가 많은데, 상위 기업 문서에서는 이러한 하위 지점/지부의 지리적 정보가 제공되지 않으므로 연결하지 않는다.
 - 예제
 - (건물) 다이소 춘천명동점
 - <다이소 춘천명동점: AF_BUILDING>은 4층까지 있어요.
 - (개체명) 다이소 춘천명동점: AF_BUILDING / (위키피디아) NONE (연결 X)
- 단, '국립중앙박물관', '국립국어원'과 같이 해당 기관에 대해 하나의 건물이 존재하며, 문서 안에서 관련 정보로서 건물에 대한 내용도 다루고 있는 경우에는 AF로 태깅되어 있더라도 해당 기관에 대한 위키피디아 문서를 연결한다.

(8) 동음이의 관계의 제품명

- 자동차, 카메라, 컴퓨터 등의 전자/기계 제품 등은 서로 다른 제품임에도 제품명이 동일한 경우가 있다. 예를 들어, 'S90'은 볼보의 자동차 제품명이기도 하지만 금영의 노래방 기계 제품명이기도 하다. 위키피디아는 이러한 제품명의 특성을 고려하여, 제품명의 경우 제조사와 제품명을 문서의 제목으로 하여 중의성을 해소하고 있다. 따라서 대상 문서의 맥락을 통해 개체명이 어떤 기업의 어떤 제품인지 특정할 수 있는 경우, 해당 제품에 대한 위키피디아 문서(제조사 + 제품명)를 연결한다.
 - 예제
 - (제품) S90
 - 다음 차는 볼보로 살까 해요. <S90: AF_TRANSPORT>이 제일 끌리긴 하는데.
 - (개체명) S90: AF_TRANSPORT → (위키피디아) 볼보 S90

개체명	위키피디아																
S90 AF_TRANSPORT	<p>볼보 S90 자 14개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>볼보 S90(Volvo S90)은 스웨덴의 고급 자동차 브랜드인 볼보에서 공개한 전륜구동 준대형 세단이다.</p> <p>1세대 [편집]</p> <p>2세대 [편집]</p> <p>2016년에 북미 국제 모터쇼에서 공개되어 볼보 S80을 대체한다. 볼보 세단 모델 최초로 SPA 플랫폼 아키텍처를 적용하였다. 대한민국에는 2016년 9월에 출시되었다. 엔진은 볼보에서 새로 개발한 Drive-E 파워 트레인 모듈 엔진은 가솔린 엔진과 디젤엔진에 둘다 사용된다. 하이브리드 엔진은(7단) 슈퍼차저 & 태보차저 + 전기모터를 합쳐 400마력을 내고 최대토크 653kgm를 낸다. 가솔린 엔진은 트윈차저(수퍼차저와 터보차저를 서로 합친 것)로 구성된다. 모든 볼보 차량 엔진은 4기통 2리터 엔진으로 통합되었다. 국내에 수입되는 모델은 D5 2.0 235마력 T6 2.0 320마력이다. 이후 T8을 기반으로 공력을 더 늘리고, 횡방향에 독립 리클라이닝 시트, 간접 태미플, 냉장고, 유리컵 등의 고급 버전의 액세서리가 출시되었다. 원래 스웨덴에서 생산하던 이후 중국 닝보에서 생산한다. 2020년에 페이스리프트를 거쳐 대일드 하이브리드가 추가되었다. 한국에서는 페이스리프트부터 일반 내연기관 가솔린 모델을 들여오지 않고, 대일드 하이브리드와 플러그인 하이브리드만 들여오며, 공력이 액셀러슨과 같은 크기로 커졌다. 2500여대의 사전계약을 달성했다.</p> <p>볼보 S90/V90</p> <table border="1"> <tr><td>차종</td><td>준대형차</td></tr> <tr><td>제조사</td><td>볼보 자동차</td></tr> <tr><td>생산년도</td><td>2016년~현재</td></tr> <tr><td>조합</td><td> 중국 하이퉁강령 자동차(S90) 대한민국 스웨덴(V90)</td></tr> <tr><td>차체형식</td><td>세단 스테이션 왜건</td></tr> <tr><td>구동방식</td><td>전륜구동 4륜구동</td></tr> <tr><td>차대</td><td>모노코크</td></tr> <tr><td>엔진</td><td>가솔린 디젤</td></tr> </table> 	차종	준대형차	제조사	볼보 자동차	생산년도	2016년~현재	조합	 중국 하이퉁강령 자동차(S90) 대한민국 스웨덴(V90)	차체형식	세단 스테이션 왜건	구동방식	전륜구동 4륜구동	차대	모노코크	엔진	가솔린 디젤
차종	준대형차																
제조사	볼보 자동차																
생산년도	2016년~현재																
조합	 중국 하이퉁강령 자동차(S90) 대한민국 스웨덴(V90)																
차체형식	세단 스테이션 왜건																
구동방식	전륜구동 4륜구동																
차대	모노코크																
엔진	가솔린 디젤																

- 제품에 대한 문서가 위키피디아에 존재하지 않는 경우, ‘없음’에 해당하는 정보를 표시한다. ‘제조사’ 문서를 연결하지 않는다.

○ 예제

- (제품) XC40
- 요즘 인기몰이 중인 볼보의 <XC40: AF_TRANSPORT>.
- (개체명) XC40: AF_TRANSPORT / (위키피디아) XC40 NONE (연결 X)

2.3.3. ORGANIZATION

(1) 하위 브랜드와 상위 기업

- 개체명 ‘파리바게트’를 위키피디아에서 검색하면 주식회사 ‘파리크라상’ 문서로 넘어간다. ‘파리바게트’는 파리크라상이 보유하고 있는 브랜드 중 하나이기 는 하나, ‘파리바게트’와 ‘파리크라상’을 정확하게 동일한 대상으로 보기는 어렵다. 따라서 특정 기업이 보유하고 있는 브랜드가 위키피디아에 별도의 문서로 존재하지 않고 상위 기업 문서로 넘어가는 경우, 개체명 하위 브랜드에 위키피디아 상위 기업 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (기업) 파리바게트
- <파리바게트: OGG_FOOD> 봄 신제품 드셔 보셨나요?
- (개체명) 파리바게트: OGG_FOOD / (위키피디아) NONE (연결 X)

(2) 기업명과 상품명

- 개체명 태깅 시 기업명으로 해당 기업이 생산하는 상품을 가리키는 것이 맥락을 통해 명확하게 드러나는 경우에는 OGG_ECONOMY가 아닌 AFW_OTHER_PRODUCTS로 태

강하도록 되어 있다. 그러나 개체명과 위키피디아 문서의 연결은 언어 표현과 실제 세계를 연결하는 것인데, 언어 표현의 본질적인 의미가 아닌 맥락을 통해 새롭게 획득한 의미를 실제 세계와 연결하면 문제가 될 수 있다. 즉, 언어의 본질적 의미와 무관한 실제 세계의 대상이 연결될 수 있는 것이다. 따라서 개체명이 기업의 명칭임에도 불구하고 AFW_OTHER_PRODUCTS로 태깅되어 있는 경우, 기업명 혹은 상품명 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (상품) 애플
- 아이폰 12랑 갤럭시 S21 고민하다가 결국 <애플: AFW_OTHER_PRODUCTS> 삼.
- (개체명) 애플: AFW_OTHER_PRODUCTS / (위키피디아) NONE (연결 X)

(3) 통폐합 혹은 개편되어 명칭이 변경된 기관이나 기업

- 기관이나 기업이 통폐합 혹은 개편되어 명칭이 변경된 경우가 있다. 예를 들어, ‘한국소프트웨어진흥원’은 공공기관 선진화 2차 계획에 따라 지식경제부(현 과학기술정보통신부 및 기획재정부) 산하 IT 관련 3개 기관(정보통신연구진흥원, 한국소프트웨어진흥원, 한국전자거래진흥원)을 통합하여 ‘정보통신산업진흥원’으로 통합·개편되었다. ‘한국소프트웨어진흥원’ 자체가 완전히 없어진 것은 아니나, 단순히 명칭만 변경된 것이 아니라 기관의 조직이나 구성도 변화되었으므로 ‘한국소프트웨어진흥원’과 ‘정보통신산업진흥원’은 지시 대상이 완전하게 일치한다고 볼 수 없다. 따라서 기관이나 기업이 통폐합 혹은 개편되어 명칭이 변경된 경우에는 변경되기 이전의 개체명에 변경된 이후의 명칭을 제목으로 하는 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (기관) 한국소프트웨어진흥원
- <한국소프트웨어진흥원: OGG_SCIENCE>에서 Winter of Code 2008을 주최했다.
- (개체명) 한국소프트웨어진흥원: OGG_SCIENCE / (위키피디아) NONE (연결 X)

- 단, 동일한 기관이나 기업인데 단순히 명칭만 변경된 경우에는 연결한다.

(4) 통합, 분리 혹은 개편되어 명칭이 변경된 정당

- 정당 또한 통합이나 분리, 개편되어 명칭이 변경된 경우가 있다. 예를 들어, ‘국민의힘’은 ‘미래통합당’이 당명을 변경한 것이며, ‘미래통합당’은 자유한국당과 새로운보수당 및 미래를향한전진4.0 등의 정당들이 통합하여 창당된 정당이다. 한국의 정당은 다른 정당과의 통합이나 기존 정당에서의 분리, 당명의 변경 등이 잦으며 명칭 변경 전과 후의 지시 대상이 완전

하게 일치한다고 볼 수 없다. 그러므로 개체명과 명확하게 일치하는 정당 문서가 존재하는 경우에만 연결하고, 변경된 최신의 당명 문서로 자동으로 넘어가는 경우에는 연결하지 않는다.

○ 예제

- (정당) 한나라당
- 1997년 11월에 <한나라당: OGG_POLITICS>이 출범하였다.
- (개체명) 한나라당: OGG_POLITICS / (위키피디아) NONE (연결 X)

- 정당의 명칭은 동음이의어가 많으므로 연결에 주의를 요한다. 위키피디아는 동음이의 정당의 경우 명칭 뒤에 창당 연도를 표시하여 중의성을 해소하고 있다. 개체명 분석 대상 맥락에서 연도를 특정할 수 있는 경우 정당의 명칭과 연도가 일치하는 문서를 찾아 연결하고, 대상 맥락에서 연도를 특정할 수 없으면 '없음'에 해당하는 정보를 표시한다.

○ 예제

- (정당) 한나라당
- 2013년 이태희가 창당한 <한나라당: OGG_POLITICS>.
- (개체명) 한나라당: OGG_POLITICS → (위키피디아) 한나라당 (2013)

개체명	위키피디아
한나라당 OGG_POLITICS	<p>한나라당 (2013년) 위키백과, 우리 모두의 백과사전.</p> <p>한나라당은 2013년 창당한 대한민국의 정당이다.</p> <p>역사 [편집]</p> <p>2012년 당시 집권당이었던 한나라당이 새누리당으로 당명을 바꾼 뒤, 영남신당자유평화당이 '한나라당'으로 이름을 바꿨지만, 4월 총선에서 득표를 받지 못해 등록취소를 당했다. 등록취소를 당한 한나라당의 대표였던 이태희가 2013년 4월 새한나라당을 창당해서, 2014년 2월 한나라당으로 이름을 바꿨다.</p> <p>외부 링크 [편집]</p> <ul style="list-style-type: none"> 한나라당 홈페이지 

- 단, 정당이 통합이나 분리 혹은 개편되면서 명칭도 함께 변경된 것이 아니라 단순히 명칭만 변경된 경우에는 연결한다.

(5) 국가대표 스포츠팀을 의미하는 국가명

- 분석 대상 문서에서 '국가명'으로 '국가대표 스포츠팀'을 지시하고 있는 경우, 즉 국가명이 LCP_COUNTRY가 아니라 OGG_SPORTS로 태깅되어 있는 경우에는 위키피디아의 '국가' 문서가 아닌 '국가대표팀' 문서를 연결한다.

○ 예제

- (국가대표팀) 독일
- 오늘 밤, <독일: OGG_SPORTS> 대 프랑스의 월드컵 최종 예선 경기가 중계됩니다.
- (개체명) 독일: OGG_SPORTS → (위키피디아) 독일 축구 국가대표팀

■ (개체명) 한국: OGG_SPORTS → (위키피디아) 대한민국 U-20 축구 국가대표팀

개체명	위키피디아
한국 OGG_SPORTS	<p>대한민국 U-20 축구 국가대표팀</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>대한민국 U-20 축구 국가대표팀은 대한민국을 대표하는 20세 이하 축구 국가대표팀이다. FIFA U-20 월드컵 본선에는 14번 출전하여 이 가운데 2019년 대회에서 준우승이라는 역대 최고의 성과를 올렸고 AFC U-19 축구 선수권 대회 본선에는 37번 출전하여 12번의 우승컵을 들어올렸고 준우승도 5번 차지했으며 4강에도 9번 진출하는 등 각종 청소년 국제 대회에서 성과를 거두었다.</p> <p>대회별 주요 성적 [편집]</p> <p>대륙 선수권 대회 [편집]</p> <ul style="list-style-type: none"> AFC U-19 축구 선수권 대회 <ul style="list-style-type: none"> 우승 (12, 최다 기록): 1959, 1960, 1963, 1978, 1980, 1982, 1990, 1996, 1998, 2002, 2004, 2012 준우승 (5): 1962, 1971, 1972, 1992, 2018 3위 (8): 1968, 1970, 1973, 1974, 1976, 2006, 2008, 2010 4위 (1): 1964 <p>세계 선수권 대회 [편집]</p> <ul style="list-style-type: none"> FIFA U-20 월드컵 <ul style="list-style-type: none"> 2위 (1): 2019 4위 (1): 1983 

- 해당 스포츠 종목에 ‘유소년 국가대표팀’ 문서가 별도로 존재하지 않는 경우 NONE으로 처리한다.
- 분석 대상 문서에서 국가명이 OGG_SPORTS로 태깅되어 있기는 하나, 그 종목을 특정할 수 없는 경우, 위키피디아 문서를 연결하지 않고 NONE으로 처리한다.

○ 예제

- (국가대표팀) 이탈리아
- 이탈리아 3 : 브라질 2
- (개체명) 이탈리아: OGG_SPORTS / (위키피디아) NONE

(6) 보통명사 ‘정부’와 특정 국가나 정권의 ‘정부’

- 대상 문서에서 ‘정부’가 특정 국가의 정부나 특정 정권을 가리킨다고 하더라도 이를 고려하지 않고 문서에서 쓰인 개체명의 언어 표현과 일치하는 위키피디아 ‘정부’ 문서를 연결한다.

○ 예제

- (기관) 정부
- 본교가 <정부: OGG_POLITICS>가 공인하는 ‘잘 가르치는 대학’으로 선정됐다.
- (개체명) 정부: OGG_POLITICS → (위키피디아) 정부

개체명	위키피디아
정부 OGG_POLITICS	<p>정부</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>다른 뜻에 대해서는 정부 (동음이의) 문서를 참고하십시오.</p> <p>행정부는 여기로 연결됩니다. 국가 내부의 행정을 담당하는 좁은 부서에 대해서는 내무부 문서를 참조하십시오.</p> <p>정부(政府, Government)는 어느 지역 또는 국가를 통치하는 기관을 말한다. 넓은 의미로는 중앙정부와 지방정부를 모두 포함하며, 좁게는 중앙정부의 행정부만을 의미하기도 한다.</p> <p>정부는 대통령제와 의원내각제를 합한 국가라면 내각도 포함한다. 그리고 만약 일당제 독재국가라면 당도 포함한다. 대한민국의 경우 좁은 의미로 사용된다. 대통령과 행정부로 구성되어, 행정권은 대통령을 수반으로 하는 정부에 속한다.</p> 

- 국립국어원의 <2020년 개체명 분석 말뭉치 구축 지침>에 따르면 ‘대한민국 정부’는 <대한민국: LCP_COUNTRY> <정부: OGG_POLITICS>로, ‘김영삼 정부’는 <김영삼: PS_NAME> <정부: OGG_POLITICS>로 분리하여 태깅하도록 되어 있다. ‘정부’ 앞에 국가나 정권에 대한 표현이 이미 제시되어 있으므로 이때의 ‘정부’도 위키피디아의 ‘대한민국 정부’, ‘김영삼 정부’ 문서를 연결하는 것이 아니라 보통명사 ‘정부’ 문서를 연결한다.

○ 예제

- (기관) 대한민국 정부
- 대한민국 <정부: OGG_POLITICS>, 태국-미얀마에 난민캠프 구축 논의
- (개체명) 정부: OGG_POLITICS → (위키피디아) 정부

- 단, 국가명이 LC가 아닌 OGG로 태깅되어 있는 경우에는 위키피디아의 ‘국가명 + 정부’ 문서를 연결한다.

○ 예제

- (기관) 프랑스
- <프랑스: OGG_POLITICS>도 '방송 장악' 논란
- (개체명) 프랑스: OGG_POLITICS → (위키피디아) 프랑스 정부

개체명	위키피디아
프랑스 OGG_POLITICS	

2.3.4. LOCATION

(1) 다의적 지명

- 기본적으로 대상 문서에서 태깅된 개체명의 태그와 동일한 주제의 위키피디아 문서를 연결한다. 예를 들어, 제주도는 행정구역으로서의 ‘도’를 의미하기도 하고 ‘섬’을 의미하기도 하는데, 문서에서 LCP_PROVINCE로 태깅되어 있으면 행정구역인 제주도를 주제로 하는 위키피디아 문서를 연결하고, LCG_ISLAND로 태깅되어 있으면 섬인 제주도를 주제로 하는 위키피디아 문서를 연결한다.

○ 예제

- (행정구역) 제주도
- <제주도: LCP_PROVINCE>의 코로나 확진자가 다시 두 자릿수로 늘었다.
- (개체명) 제주도: LCP_PROVINCE → (위키피디아) 제주특별자치도

개체명	위키피디아
<p>제주도</p> <p>LCP_PROVINCE</p>	<p>제주특별자치도</p> <p>위키택과, 우리 모두의 백과사전.</p> <p>다른 문서와 다뤄서는 제주 문서를 참고하십시오.</p> <p>제주특별자치도(濟州特別自治道)는 대한민국의 제주도(濟州島)와 부속 섬들을 관할하는 특별자치도이다. 대한민국에서 가장 큰 섬인 제주도(濟州島) 본섬을 비롯하여 다라도, 우도, 주자군도 등을 포함한 유인도 8개, 무인도 55개로 구성되어 있다. 한라산 일대의 아름다운 자연 경관으로 관광업, 특히 배스업과 어업이 발달하였고, 제주 화산섬과 용암 동굴은 유네스코 세계 자연 유산에 등재되었다. 조선시대 전라도에 속했으나, 이순절 통치 하에 전라남도에서 분리되었고, 1946년 전라남도에서 분리되어 제주도(濟州道)가 되었다.</p> <p>2006년부터는 특별자치도로 승격되어 자치경찰의 실시, 교육자치권의 확대, 일부 중앙관청의 이양, 자치입법권, 자치재정권의 부여 등 지역적 자치권이 고도화되었다. 하위에 기초지방자치단체를 두지 않은 단층제 광역지방자치단체이며, 남북으로 두 개의 행정시(市)인 제주시와 서귀포시가 설치되어 있다.</p> 

○ 예제

- (섬) 제주도
- <제주도: LCG_ISLAND>는 대한민국에서 가장 큰 화산섬이다.
- (개체명) 제주도: LCG_ISLAND → (위키피디아) 제주도

개체명	위키피디아
<p>제주도</p> <p>LCG_ISLAND</p>	<p>제주도</p> <p>위키택과, 우리 모두의 백과사전.</p> <p>다른 문서와 다뤄서는 제주도 (동음이의) 문서를 참고하십시오.</p> <p>제주도(濟州島, 영어: Jeju-do)는 대한민국의 남해에 위치한 섬이자 대한민국에서 가장 큰 섬이다. 또한 제주도는 행정구역상 제주특별자치도에 속한다. 섬의 면적은 1833.2 km²로 이는 남한 면적의 1.83%에 해당한다. 2020년 주민등록 인구는 약 67만 명으로, 이는 대한민국의 섬 중에서 가장 인구가 많다.</p> <p>또한 제주도는 중앙의 한라산을 중심으로 완만한 경사를 이루어 동서 73 km, 남북 31km의 타원형 모양을 하고 있다. 일주도로 길이는 181km, 해안선은 258km이다. 제주도의 북쪽 끝은 김녕해수욕장이고, 남쪽 끝은 송악산이며, 서쪽 끝은 수월봉, 동쪽 끝은 성산 일출봉이다. 제주도는 온대 기후에 속하여, 겨울에도 거의 영하로 떨어지지 않고 영상을 유지한다. 황해와 동해, 동중국해 경계에 자리 잡아 대한민국의 정치·경제는 물론 군사상으로도 중요한 지위에 있다.</p> <p>또한 제주도는 2002년 12월 16일 유네스코가 기후 및 생물 다양성의 생태계적 가치를 인정하여 제주도를 생물권보전지역을 지정하였다. 가과도 뿐만 아니라 한라산, 성산 일출봉, 거문오름 등암등굴, 거가 화물문화관광 생태 등의 가치와 중요성을 인정받아 2007년 6월, 제주 화산섬과 용암 동굴이라는 이름으로 세계자연유산에 등록되었다. 2010년 10월 3일 유네스코 세계지질공원네트워크(IGGN)가 제주도 지역에 세계지질공원을 인증했다.</p> 

(2) 동일한 상위 행정구역명과 하위 행정구역명

- 상위의 행정구역명과 하위의 행정구역명이 동일한데 개체명 태그도 동일한 경우, 대상 문서의 맥락을 통해 어떤 행정구역인지 판별하고, 그에 맞는 위키피디아 문서를 연결한다.

○ 예제

- (행정구역) 창녕
- <창녕: LCP_COUNTY>, 대구 39도를 찍은 날
- (개체명) 창녕: LCP_COUNTY → (위키피디아) 창녕군

개체명	위키피디아
<p>창녕</p> <p>LCP_COUNTY</p>	<p>창녕군</p> <p>위키택과, 우리 모두의 백과사전.</p> <p> 다른 뜻에 대해서는 창녕군 (동음이의) 문서를 참고하십시오.</p> <p>창녕군(昌寧郡)은 대한민국 경상남도 중북부에 있는 군이다. 군청 소재지는 창녕읍이고, 행정구역은 2읍 12면이다.</p>

- ‘창녕’과 같이 나열된 ‘대구’가 광역시이므로, 앞의 ‘창녕’ 또한 읍, 면, 동, 리 이상의 행정구역으로 판단된다. 따라서 ‘창녕읍’이 아니라 ‘창녕군’ 위키피디아 문서를 연결한다.
- 맥락을 통해 대상 개체명의 행정구역을 명확하게 판별할 수 없는 경우, 상위의 행정구역에 대한 위키피디아 문서를 우선적으로 연결한다.
 - 예제
 - (행정구역) 합천
 - <합천: LCP_COUNTY> 왔다.
 - (개체명) 합천: LCP_COUNTY → (위키피디아) 합천군

개체명	위키피디아
<p>합천</p> <p>LCP_COUNTY</p>	<p>합천군</p> <p>위키택과, 우리 모두의 백과사전.</p> <p> 합천은 여기로 연결됩니다. 다른 뜻에 대해서는 합천 (동음이의) 문서를 참고하십시오.</p> <p> 다른 뜻에 대해서는 합천군 (동음이의) 문서를 참고하십시오.</p> <p>합천군(陜川郡)은 대한민국 경상남도 서북부에 있는 군이다. 군청 소재지는 합천읍이고, 행정구역은 1읍 16면이다. 가야면에는 세계문화유산인 장경판전과 세계기록유산인 말안대장경이 있는 해인사가 있다.</p>

- 동일한 명칭을 가지는 상위 행정구역과 하위 행정구역은 전체-부분 관계이므로 상위의 행정구역명을 연결하여도 지시 대상이 완전히 달라지는 것은 아니다. 또한 ‘군’이 나 ‘읍’과 같은 행정구역 단위를 발화하지 않고 지명만 발화하는 경우, 일반적으로 ‘군’ 단위를 가리키는 것이 한국어 지명 사용상의 기본값이라고 할 수 있다. 따라서 맥락을 통해 대상 개체명의 행정구역을 명확하게 판별할 수 없는 경우, 상위의 행정구역에 대한 위키피디아 문서를 연결한다.

(3) 동명의 지명

- ‘서면’, ‘교동’, ‘우도’ 등과 같이 서로 다른 지역에 같은 지명이 존재하며 개체명 태그도 동일한데, 대상 문서의 맥락을 통해서도 개체명이 가리키는 지역이 어디인지 명확하게 판별할 수 없는 경우, 위키피디아 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (행정구역) 광주

- 여기서 <광주: LCP_CITY>까지 3시간 걸려.
- (개체명) 광주: LCP_CITY / (위키피디아) NONE (연결 X)

○ 위의 문장 ‘여기서 광주까지 3시간 걸려.’에서 개체명 ‘광주’가 전라남도의 ‘광주광역시’인지 경기도의 ‘광주시’인지 맥락으로 판별할 수 없다. 임의로 개체명이 가리키는 대상을 특정할 수는 없으므로 위키피디아 문서를 연결하지 않는다.

(4) 전체와 부분을 가리키는 지명, 해당 지명에 대한 위키피디아 문서가 각각 존재하는 경우

● 한국 / 대한민국 / 조선민주주의인민공화국

○ 일상생활에서 ‘한국’이라는 용어를 자주 사용하는데, 한국 사회에서 일반적으로 사용되는 ‘한국’은 ‘북한’을 포함하지 않는 ‘대한민국’만을 가리키는 경우가 많다. 위키피디아에는 ‘한국’, ‘대한민국’, ‘조선민주주의인민공화국’ 문서가 각각 존재하며, ‘한국’ 문서는 ‘대한민국’과 ‘조선민주주의인민공화국’을 모두 포함한 ‘한국’에 대한 문서이다. 따라서 일반적으로 남한을 이르는 말로 ‘한국’이 쓰였을 때는 위키피디아 ‘한국’ 문서를 연결하는 것이 아니라 ‘대한민국’ 문서를 연결하여야 한다. 남한과 북한을 모두 아우르는 의미로 ‘한국’이 쓰였을 경우에만 위키피디아 ‘한국’ 문서를 연결한다.

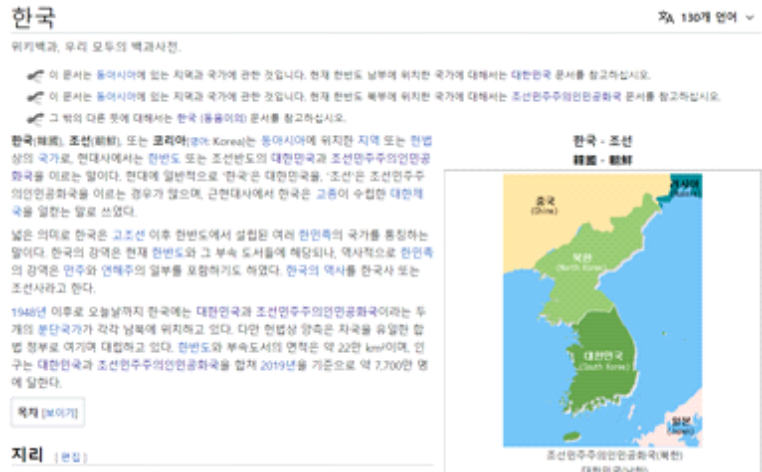
○ 예제

- (국가) 한국
- <한국: LCP_COUNTRY> 엄마들은 누구나 이런 제품 원하잖아요.
- (개체명) 한국: LCP_COUNTRY → (위키피디아) 대한민국

개체명	위키피디아
한국 LCP_COUNTRY	

○ 예제

- (국가) 한국
- 분단 이전 <한국: LCP_COUNTRY>의 역사 탐구
- (개체명) 한국: LCP_COUNTRY → (위키피디아) 한국

개체명	위키피디아
<p>한국</p> <p>LCP_COUNTRY</p>	

● 중국 / 중화인민공화국 / 중화민국(대만)

- 위키피디아 ‘중국’ 문서에서는 “중국(중국어 간체자: 中国, 정체자: 中國, 병음: Zhōngguó 중궈[*], 영어: China 차이나[*])은 중화인민공화국 또는 중화민국을 일컫는 단어이다”로 중국을 정의하고 있으며, “사람들마다 개인별로 차이가 있으나 중화인민공화국만을 지칭하는 경우가 중화민국을 지칭하는 경우보다 훨씬 많아졌고, 문맥에 따라서는 둘을 포괄하여 지칭되기도 한다.”라고 그 범위를 설명하고 있다. 이에 따라 개체명과 위키피디아 문서를 연결할 때 중화인민공화국과 중화민국을 포괄하는 의미의 ‘중국’, 중화인민공화국과 중화민국으로 나뉘기 전의 ‘중국’을 의미하는 경우에만 위키피디아 ‘중국’ 문서를 연결한다. 중국 본토를 의미하는 경우에는 ‘중화인민공화국’ 문서를, 자유중국(대만)을 의미하는 경우에는 ‘중화민국’ 문서를 연결한다. (‘대만’의 경우, 섬을 의미할 때에는 ‘타이완’ 문서를 연결한다.)

○ 예제

- (국가) 중국
- <중국: LCP_COUNTRY> – 대만 파워에 신남방정책 성과 흔들
- (개체명) 중국: LCP_COUNTRY → (위키피디아) 중화인민공화국

개체명	위키피디아
중국 LCP_COUNTRY	<p>중화인민공화국 291개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>중화인민공화국(중국어 간체자: 中华人民共和国, 정체자: 中華人民共和國, 병음: Zhōnghuá Rénmín Gònghéguó 중화인민공화국)은 동아시아에 속하는 국가이다. 줄여서 중국(중국어 간체: 中国, 정체: 中國, 병음: Zhōngguó 중궈)이라고도 한다.</p> <p>수도는 베이징이며, 최대 도시는 상하이이다. 이밖에 중화인민공화국의 주요 도시로는 광저우, 난징, 지안, 무저우, 칭다오, 충칭 등이 있다.</p> <p>근대 중국에서는 1911년 신해혁명으로 청나라가 무너지고 국민당 정부가 들어섰으나, 정치적 혼란이 이어지면서 각지에서 군벌정부가 권력을 휘두르는 군벌 시대로 돌입하게 되었다. 한편 5.4 운동의 반제국주의 감정을 중심으로 창설된 중국 공산당은 1920년 이후 소련의 지원을 받으며 세력을 급격히 키웠고, 국민정부와 계속해서 내전을 벌였다. 중일 전쟁 이후 중국 대륙의 주도권을 두고 1946년 제2차 국공 내전이 발발하였고, 그 결과 타이완 등 일부 지역을 제외한 영역이 공산당 측으로 넘어가 1949년 다오회동을 통해 국가주석으로 하여 베이징에서 중화인민공화국의 성립이 선포되었다.</p> 

○ 예제

- (국가) 중국
- 현재 <중국: LCP_COUNTRY>은 여러 가지 정치상 문제로 말미암아 완벽한 통일국가를 이루지 못하고 있으며, 다수의 정부가 존재한다.
- (개체명) 중국: LCP_COUNTRY → (위키피디아) 중국

개체명	위키피디아
중국 LCP_COUNTRY	<p>중국 78개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>1949년에 건국된 중국 대륙에 위치한 국가에 대해서는 중화인민공화국 문서를, 1912년에 건국된 타이완 지부에 위치한 국가에 대해서는 중화민국 문서를 참조하십시오. 다른 뜻은 중화민국 (중국어) 또는 중국 (중국어) 문서를 참조하십시오.</p> <p>중국(중국어 간체자: 中国, 정체자: 中國, 병음: Zhōngguó 중궈), 영어: China 타이완(台)은 중화인민공화국 또는 중화민국을 일컫는 단어이다. 본디 중국은 청나라 때까지 한 국가였으나, 1949년에 국공내전에서 승자가 갈라면서 두 국가로 된 이래 현재까지 중화인민공화국과 중화민국, 즉 두 개의 중국이 존재하고 있다. 그러나 각국은 하나의 중국을 주장하며 상대를 인정하지 않고 있다.</p> <p>특징 (해당)</p> <p>중국의 범위 (편집)</p> <p>사람마다 개인별로 차이가 있으나 중화인민공화국을 지칭하는 경우가 중화민국을 지칭하는 경우보다 훨씬 많아졌고, 문맥에 따라서는 둘을 포괄하여 지칭되기도 한다. 때로는 중화권과 동일시되기도 한다. 1980년대까지만 하더라도 중화인민공화국과 수교하지 않은 대한민국에서는 '자유중국'이라고 하여 중화민국을 중국이라 불렀으나, 수교 이후부터는 점차 중화인민공화국을 중국으로 부르기 시작한 것이다.</p> <p>역사 (편집)</p> 

● 강원도 / 강원도 (남) / 강원도 (북)

- ‘강원도’는 한국이 ‘대한민국’과 ‘조선민주주의인민공화국’으로 분단되면서 둘로 나뉘었다. ‘대한민국’에도 ‘강원도’가 있고 ‘조선민주주의인민공화국’에도 ‘강원도’가 있다. 이에 따라 위키피디아에는 분단 이전의 ‘강원도’, 대한민국의 ‘강원도’, 조선민주주의인민공화국의 ‘강원도’ 문서가 각각 존재한다. 따라서 맥락을 통해 ‘강원도’가 강원도 전체 지역을 의미하는지, 남한의 강원도를 의미하는지, 북한의 강원도를 의미하는지 판별하여 강원도 전체 지역을 의미할 때는 ‘강원도’, 대한민국의 강원도를 의미할 때는 ‘강원도 (남)’, 조선민주주의인민공화국의 강원도를 의미할 때는 ‘강원도 (북)’ 문서를 연결한다. 맥락으로 판단할 수 없는 경우에는 ‘강원도’ 문서를 우선적으로 연결한다.

○ 예제

- (지명) 강원도
- 이번 여름 휴가는 <강원도: LCP_PROVINCE> 양양으로!
- (개체명) 강원도: LCP_PROVINCE → (위키피디아) 강원도 (남)

개체명	위키피디아
강원도 LCP_PROVINCE	<p>강원도 (남) 총 53개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>강원도(江原道)는 대한민국의 북동부에 있는 도이다. 동쪽은 동해, 서쪽은 경기도, 남쪽은 충청북도·강원북도와 접하며, 북쪽은 북한의 강원도와 경계를 이루고 있다. 지리적으로는 태백산맥을 기준으로 동쪽으로 강릉시, 속초시, 동해시, 삼척시, 태백시 등 영동 지방과 서쪽으로 춘천시, 원주시 등 영서 지방으로 구분한다. 강원도청은 춘천시에 위치하고 있으며, 강릉시에 강원도청의 출장소인 강원도원동해본부가 설치되어 있다. 행정 구역은 7시 11군이다.</p> <p>특차 [보여기]</p> 

○ 예제

- (지명) 강원도
- 금강산은 조선민주주의인민공화국 <강원도: LCP_PROVINCE> 금강군 · 고성군 · 통천군에 걸쳐 있다.
- (개체명) 강원도: LCP_PROVINCE → (위키피디아) 강원도 (북)

개체명	위키피디아
강원도 LCP_PROVINCE	<p>강원도 (북) 총 48개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>강원도(江原道)는 조선민주주의인민공화국 남동부에 있는 도이다. 1948년 9월에 함경남도에서 공산시, 안변군, 문천군이 강원도에 편입되었고, 도청 소재지는 공산시이다. 당위용회 위용장은 박정남이고, 연탄 위용회 위용장은 현성준이다.</p> <p>특차 [보여기]</p> <p>지리 [편집]</p> <p>동쪽으로는 동해, 서쪽으로는 황해북도, 북쪽으로는 함경남도 및 평안남도, 남쪽으로는 군사분계선을 경계로 대한민국의 강원도 및 경기도 연천군과 접하고 있다.</p> <p>도의 대부분을 태백산맥이 통과하고 있고, 가장 높고 유명한 산은 금강산이다. 조선민주주의인민공화국의 강원도와 대한민국의 강원도를 합쳐 관동 지방으로 부른다. 태백산맥의 서쪽을 영서, 동쪽을 영동이라 부른다.</p> 

○ 예제

- (지명) 강원도
- <강원도: LCP_PROVINCE>는 강릉의 ‘강’과 원주의 ‘원’을 합해 만들어진 지명이다.
- (개체명) 강원도: LCP_PROVINCE → (위키피디아) 강원도

개체명	위키피디아
강원도 LCP_PROVINCE	<p>강원도 2A 19개 언어</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>강원도는 한국의 분단 이전에 존재했던 강원도에 관한 것입니다. 현재의 강원도에 대해서는 강원도 (남)과 강원도 (북) 문서를, 다른 뜻에 대해서는 강원도 (동음이의) 문서를 참조하십시오.</p> <p>강원은 여기로 연결됩니다. 불교의 경전을 배우는 학교와 같은 교육기관인 강원(講院)에 대해서는 종당 문서를 참조하십시오.</p> <p>강원도(江原道)는 관동 지방(關東地方)에 해당하는 지역이다. 1895년에 설치되었고 강원도라는 이름은 도의 주요 도시인 강릉과 감영 소재지였던 원주에서 따 온 것이다.</p> <p>목록 [보이기]</p> <p>위치 [편집]</p> <p>강원도는 한반도 중앙부의 동쪽에 태백산맥을 중심으로 크게 구분되어 있다. 위도상으로는 북위 37도 02분에서 38도 17분에 걸쳐고 경도상으로는 동경 127도 05분에서 129도 22분에 걸쳐 있으며 북위 38도선은 본 도의 거의 중앙부를 통과하고 휴전선은 고성군 한대면 북위 38도 35분 근처에서 서남하여에 함보령, 문동리 및 김화읍의 북방을 연결하는 북위 38도 20분선 부근에서 145km에 걸쳐 그어져 있다.</p> <p>강원도 동서의 길이는 약 150km, 남북은 약 243km에 달하며, 동쪽은 약 314km에 걸쳐 해안선을 이루면서 이여져 있고, 서방은 황해도 및 경기도와 접해 있고 남쪽은 충청북도 및 경상북도, 북쪽은 함경남도 및 황해도와 접하여 경계를 이루고 있다.</p> 

(5) 법정동과 행정동

- 위키피디아에서 법정동을 검색하였을 때 법정동 문서는 존재하지 않고 행정동 문서로 넘어가는 경우, 개체명 ‘법정동’에 위키피디아 ‘행정동’ 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
- 행정동은 행정 운영의 편의를 위하여 설정한 행정구역으로서 주민 수의 증감에 따라 수시로 설치 또는 폐지되므로 변동성이 크며, 하나의 법정동이 분리되어 각각 다른 행정동에 포함되는 경우가 있어, 개체명이 지시하는 대상이 행정동 문서에 완전히 포함되지 않을 수 있다.
- 또한 개체명의 외연이 개체 연결 대상 표제어의 외연에 포함되는 경우도 있고 그렇지 않은 경우도 있어, 일관성을 확보하기가 어렵다.
- 따라서 법정동 검색 시 행정동 문서로 자동으로 넘어가는 경우, 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제
 - (지명) 지곡동
 - 용인 <지곡동: LCP_COUNTY>의 전원주택 단지
 - (개체명) 지곡동: LCP_COUNTY / (위키피디아) NONE (연결 X)

(6) 과거의 지명과 현재의 지명

- 과거의 지명을 위키피디아에서 검색하였을 때, 과거 지명 문서는 존재하지 않고 현재 지명 문서로 넘어가는 경우가 있다. 예를 들어, 위키피디아에서 ‘육조거리’를 검색하면 ‘세종대로’ 문서로 넘어간다. 조선시대의 ‘육조거리’와 현재의 ‘세종대로’는 해당 지역의 과거의 명칭과 현재의 명칭이나 지시하는 대상의 외연적 의미와 내포적 의미가 완전하게 일치하지 않는다. 따라서 위키피디아에 과거 지명 문서가 존재하지 않고 현재 지명 문서로 넘어가는 경우, 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.
 - 예제

- (거리) 육조거리
- 조선시대에 <육조거리: LC_OTHERS> 일대는 정치, 행정의 중심지였다.
- (개체명) 육조거리: LC_OTHERS / (위키피디아) NONE (연결 X)
- 단, '직할시'에서 '광역시'로, '리'에서 '동' 등으로 행정구역이 변경되거나 승격된 경우에는 연결한다.
 - 예제
 - (지명) 수완리
 - 여기 예전에는 <수완리: LCP_COUNTY>였잖아요.
 - (개체명) 수완리: LCP_COUNTY → (위키피디아) 수완동

개체명	위키피디아
수완리 LCP_COUNTY	<p>수완동</p> <p>위키백과, 우리 모두의 백과사전. (수완리에서 넘어옴)</p> <p>수완동(水莞洞)은 광주광역시 광산구의 행정동이다. 법정동 신가동 일부, 수완동 일부, 장덕동 일부, 흑석동 일부로 구성된다. 법정동 수완동의 나머지는 백아동에 속한다. 광주광역시 광산구 중에 제일 인구가 제일 많은 동에 속한다.^[1] 장덕도서관이 2013년 12월 27일에 개관하였다.^[2] 수완동에 북카페 놀음이 존재하며 최소 2~3가지 이상의 마을지도가 존재하고 있다.</p>

(7) 종합 쇼핑/문화 센터와 입점 브랜드

- 종합 쇼핑/문화 센터에는 다양한 상점, 문화/오락 시설이 입점되어 있다. 이러한 입점 브랜드 혹은 상점을 위키피디아에 검색하였을 때 문서가 존재하지 않고 종합 쇼핑/문화 센터로 문서가 넘어가는 경우가 있다. 입점되어 있는 상점 혹은 시설은 해당 종합 쇼핑/문화 센터 안에 포함되기는 하나, 이는 위치상의 포함 관계일 뿐이며 그 둘이 지시하는 대상은 각기 다르다. 따라서 특정 종합 쇼핑/문화 센터에 입점해 있는 상점이나 문화/오락 시설이 위키피디아에 별도의 문서로 존재하지 않고 상위의 종합 쇼핑/문화 센터 문서로 넘어가는 경우(두바이 아쿠아리움→두바이몰), 개체명 '입점 상점, 문화/오락 시설'에 위키피디아 '종합 쇼핑/문화 센터' 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.
 - 예제
 - (장소) 두바이 아쿠아리움
 - <두바이 아쿠아리움: LC_OTHERS>은 2008년 11월에 개장했다.
 - (개체명) 두바이 아쿠아리움: LC_OTHERS / (위키피디아) NONE (연결 X)

(8) 기관(OGG)이 아닌 장소(LC)로 태깅되어 있는 기관이나 기업

- 기관이나 기업이 장소(LC)의 의미를 가질 때에는 기관 혹은 기업 그 자체가 아닌 그 일대를 가리키는 것이므로 기관, 기업에 대한 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

- 예제
 - (지명) 한남대
 - <한남대: LC_OTHERS> 맛집 탐방.
 - (개체명) 한남대: LC_OTHERS / (위키피디아) NONE (연결 X)
- 단, '홍대 (지명)'과 같이 기관, 기업이 위치한 일대에 대한 문서가 위키피디아에 별도로 존재하는 경우에는 해당 위키피디아 문서를 연결한다.

- 예제
 - (지명) 홍대
 - 주말에 <홍대: LC_OTHERS>에서 놀았어.
 - (개체명) 홍대: LC_OTHERS → (위키피디아) 홍대 (지명)

개체명	위키피디아
홍대 LC_OTHERS	

2.3.5. DATE

(1) N월+초/말, 초순(상순)/중순/하순

- 개체명이 'N월+초/말, 초순(상순)/중순/하순'과 같은 구성인데 해당 구성 전체에 대한 위키피디아 문서가 존재하지 않는 경우, 위키피디아의 'N월' 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

- 예제
 - (달) 8월 말
 - 우리 학교는 보통 <8월 말: DT_MONTH>에 개강한다.
 - (개체명) 8월 말: DT_MONTH / (위키피디아) NONE (연결 X)

(2) 접사, 조사, 관형사 등이 결합된 표현

- '초여름', '수요일부터', '2002년경', '이번 주말' 등과 같이 날짜나 계절 표현에 접사, 조사, 관형사 등이 결합되어 있는데 해당 언어 표현과 동일한 위키피디아 문서가 존재하지 않는 경우, 위키피디아 문서를 연결하지 않고 '없음'에 해당하는 정보를 표시한다.

- 예제
 - (계절) 초여름
 - <초여름: DT_SEASON>에 입기 좋은 외투 없을까요?

- (개체명) 초여름: DT_MONTH / (위키피디아) NONE (연결 X)

○ 예제

- (날짜) 금요일까지
- <금요일까지: DT_OTHERS> 과제 제출해야 돼.
- (개체명) 금요일까지: DT_OTHERS / (위키피디아) NONE (연결 X)

(3) 국가명 + 초기, 중기, 말기

- ‘국가명 + 초기, 중기, 말기’와 같은 시기 구분 명칭은 해당 구성 전체가 위키피디아에 별도의 문서로 존재하지 않고 ‘국가’ 문서로 넘어가는 경우가 있다. 이러한 경우, 개체명 ‘국가명 + 초기, 중기, 말기’에 위키피디아 ‘국가’ 문서를 연결하지 않고 ‘없음’에 해당하는 정보를 표시한다.

○ 예제

- (시기) 고려 말기
- <고려 말기: DT_DYNASTY> 위화도 회군
- (개체명) 고려 말기: DT_OTHERS / (위키피디아) NONE (연결 X)

(4) OO 전기, 후기

- 중세, 중세 전기, 중세 후기 등의 역사 시대 구분 명칭의 경우, 대상 문서에서 연도가 제시되어 있는데 연도와 시대 구분 명칭이 맞지 않다고 하더라도, 연도는 고려하지 않고 역사 시대 구분 명칭에 맞는 위키피디아 문서를 연결한다.

○ 예제

- (시대) 중세 후기
- 1200년대 <중세 후기: DT_DYNASTY> 작품.
- (개체명) 중세 후기: DT_DYNASTY → (위키피디아) 중세 후기

개체명	위키피디아
중세 후기 DT_DYNASTY	<p>중세 후기</p> <p>위키피백과, 우리 모두의 백과사전.</p> <p>중세 후기(Late Middle Ages)는 유럽사의 시대 구분 중 하나로 14세기부터 15세기 중반까지를 의미한다. 이는 중세 초기 이후에 등장하여 이후에는 근세가 등장한다.(르네상스) 1300년 유럽 전역에서는 기근과 흑사병이 창궐하기 시작한다. 이는 유럽인들의 급감을 야기했다.^[1] 이러한 인구감소는 사회불안과 폭동을 야기했다. 프랑스와 잉글랜드의 경우 소작농들이 자주 난을 일으켰는데, 자크리의 난과 와트 타일러의 난이 대표적인 예이다. 이 뿐만 아니라 백년전쟁까지 일어났다. 게다가 기독교의 경우 서방 교회와 분열로 로마와 아비뇽에 교황이 두 명이 있게 되는 사건까지 발생한다. 이를 서양 역사가들은 중세 후기의 수난이라고 일컫는다.^[2]비록 이러한 혼란이 있더라도 14세기에는 과학과 예술에는 여러 진보적인 성과가 등장하기 시작했다. 고대 그리스 미술과 중세 중기부터 기원한 고대 로마와 라틴어 문서의 경우 이탈리아 르네상스를 이끌었다. 특히 12세기와 르네상스의 경우 십자군 전쟁 이후 그리스 학자들이 피신을 하면서 시작되었고, 이는 오스만 제국의 콘스탄티노플 정복 이후 가속화되어 비잔틴 제국의 학자들이 서방으로 피난자를 찾으면서 르네상스를 일당했다. 특히 이탈리아 반도의 경우 더욱 그랬다.^[3] 이러한 교전의 지식과 인쇄술이 결합하면서 이는 종교 개혁을 이끌었으며, 중세 말기 말엽에는 대항해시대를 열기도 했다. 오스만 제국이 성장하여 콘스탄티노플 함락을 한 1453년, 오스만 제국에 의해 동쪽 끝이 막히고 나서 신 왕조를 개척해야만 했다. 중립베스의 경우 1492년 아메리카 대륙을 발견했으며, 바스코 다 가마의 경우 인도와 아프리카의 해안을 따라 1498년 항로를 개척했다. 이 사건들은 유럽의 국가들의 경제적 군사력을 증강시키는 발전이었다. 이러한 변화는 중세와 근대 사이를 연결해주는 근세시대를 열었으며, 교전의 지식이 현대로까지 전승되게 해주는 큰 역할을 해주는 시대라고 역사가들은 평가하고 있다.</p> 

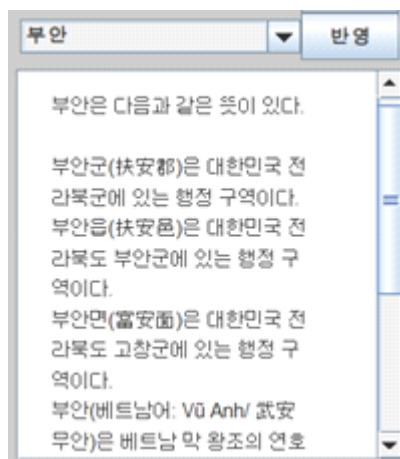
- ‘1200년대 중세 후기 작품’에서 ‘1200년대’는 역사적으로 중세 성기에 해당하나, ‘중세 후기’라는 언어 표현에 ‘중세 성기’ 위키피디아 문서를 연결할 수는 없으므로, 연도는 고려하지 않고 ‘중세 후기’ 위키피디아 문서를 연결한다.

3. 주의 사항

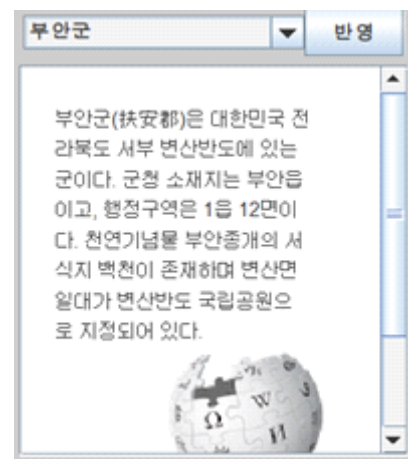
(1) 동음이의 문서의 처리

● 동음이의 페이지에 해당 문서가 있는 경우

- 도구에서 개체명을 검색하였을 때, 다음과 같은 동음이의어 문서창이 뜨는 경우가 있다. 이러한 경우, “OO에는 다음과 같은 뜻이 있다.” 아래에 나열되어 있는 단어들 중에서 개체명과 같은 대상을 지시하는 단어를 재검색한 후, 해당 문서를 개체명과 연결한다. ‘OO (동음이의)’ 페이지를 개체명에 연결해서는 안 된다.



(X)



(O)

● 동음이의 페이지에 해당 문서가 없는 경우

- 개체명은 금정구의 고분로인데, 아래와 같이 동음이의 페이지에 설명은 되어 있으나 독립적인 페이지가 존재하지 않는 경우(제목이 빨간색이거나 검은색인 경우), ‘OO (동음이의)’ 페이지를 연결하지 않고 NONE으로 처리한다.

고분로

위키백과, 우리 모두의 백과사전.

고분로는 다음의 도로가 있다.

- **고분로 (금정구)** - 부산광역시 금정구 노포동
- **고분로 (연제구)** - 부산광역시 연제구 연산5동과 연산9동을 연결하는 도로이다.



이 문서는 영칭은 같지만 대상이 다를 때에 쓰이는 **동음이의어 문서**입니다.
어떤 링크가 이 문서를 가리키고 있다면, 그 링크를 알맞게 고쳐 주세요.

분류: 동음이의어 문서 | 동음이의어 도로

- 위키피디아에 다수의 동음이의어 문서가 존재하는 경우가 있으므로, 개체명과 위키피디아 문서의 제목이 동일하다고 하여 기계적으로 연결하여서는 안 된다. 반드시 문서의 내용을 확인하고 개체명의 의미와 부합하는 내용의 문서인 경우에만 연결한다. 제목만 동일하고 내용은 개체명의 의미와 맞지 않는 경우에는 연결하지 않는다.

(2) 동일지시적 표현의 처리

● 00역, 00동, 00시 등

- 역 이름, 동 이름 등은 일상적으로 '-역', '-동'이 없이 쓰이는데, 이때 도구는 '-역', '-동'의 앞부분이 제목인 위키피디아 문서를 제공하게 된다. 위키피디아는 지하철역이나 동의 이름을 '-역', '-동'이 결합된 완전한 형태로 제공하므로 '-역', '-동'이 빠진 제목의 위키피디아 문서는 지하철역이나 동에 관한 문서가 아닐 가능성이 크다. 따라서 정식 명칭의 일부만 표현된 개체명의 경우 도구에서 제공하는 위키피디아 문서를 기계적으로 연결해서는 안 되며, 완전한 형태를 검색한 후 맞는 문서를 연결해야 한다.

● 일상적 표현

- 일반적으로 행정기관과 관청은 동일한 언어로 표현되기도 한다. 즉 '서울시청'이 행정기관의 의미로 쓰이기도 하고, 관청의 의미로 쓰이기도 하는 것이다. 개체명을 태깅할 때에는 이러한 언어 표현에 대해 맥락에 맞는 태그를 각각 부착한다. 그러나 위키피디아의 경우, 동일한 제목으로 다른 문서를 만들 수 없으므로 명확한 언어 표현으로 문서의 제목을 붙이는 경우가 많다. 따라서 행정기관인 '서울시청'은 위키피디아 문서에 '서울특별시청'이라는 제목으로 등재되어 있고, 관청인 '서울시청'은 '서울특별시청사'라는 제목으로 등재되어 있다. 이러한 위키피디아 문서의 특성을 고려해 볼 때, 태깅된 개체명과 동일한 제목의 위키피디아 문서가 존재하지 않는다고 하더라도 다른 명칭으로는 존재할 수 있으므로, 가능한 다른 언어 표현을 검색하여 개체명이 지시하는 대상과 동일한 주제의 위키피디아 문서를 연결할 수 있어야 한다.

(3) 자동으로 넘어가는 문서의 처리

- 위키피디아에서 개체명을 검색하였을 때, 해당 개체명에 대한 문서가 존재하지 않고 다른 제목의 문서로 자동으로 넘어가는 경우가 있다. 위키피디아에서 자동으로 넘어가는 문서는 동일한 대상에 대한 다른 제목의 문서일 수도 있으나, 지시 대상이 전혀 다른 문서일 수도 있다. 따라서 자동으로 넘어가는 문서의 경우, 해당 문서에서 설명하는 대상이 개체명이 지시하는 대상과 일치하는지 여부를 반드시 따져본 후 일치하면 연결하고, 일치하지 않으면 연결하지 않는다.

태깅이 잘못된 개체명의 처리

● 태그 오류

- 개체명에 태그가 잘못 할당된 경우, 위키피디아 문서를 연결하지 않고 ‘오류’에 해당하는 정보를 표시한다. 예를 들어, 맥락으로 보았을 때 PS_NAME이 분명하나, PS_PET으로 태깅되어 있다면 PS_NAME으로 수정하고 관련 위키피디아 문서를 연결하는 것이 아니라, 그대로 두고 ERROR로 처리한다.

● 단위 오류

- 개체명 태깅 단위가 잘못된 경우에도 위키피디아 문서를 연결하지 않고 ‘오류’에 해당하는 정보를 표시한다. 예를 들어, ‘스타벅스 강남점’의 경우 ‘〈스타벅스 강남점: OGG_FOOD〉’과 같이 전체를 하나의 단위로 태깅해야 하는데, ‘〈스타벅스: OGG_FOOD〉 강남점’과 같이 구성의 일부만을 태깅 단위로 설정하였다면 태깅 단위를 수정하고 위키피디아 문서를 연결하는 것이 아니라, 그대로 두고 ERROR로 처리한다.

● 태그 미부착

- 위키피디아 문서 연결 대상 대분류인 PS, AF, OGG, LC, DT에 속하는 개체명이나, 대상 문서상에서 개체명 태깅이 되어 있지 않은 경우, 작업자가 태깅을 하지 않고 위키피디아 문서도 연결하지 않는다. 대상 문서상에서 태깅이 되어 있는 개체명에 한해 위키피디아 문서를 연결한다.

(4) 오자나 오타의 처리

- 국립국어원의 <2020년 개체명 분석 말뭉치 구축 지침>에서는 ‘원시 어절에서 명백한 오타(전사상 오류, 발음 오류 등)로 보이는 경우에도 개체명 분류에 속하면 원어절 그대로 해당 개체명으로 태깅한다.’라는 오자나 오타에 대한 지침을 두고 있다. 이에 따라 위키피디아 문서를 연결할 때에도 오자나 오타임이 명확한 경우, 또한 해당 개체명이 지시하는 대상을 특정할 수 있는 경우에는 해당 대상에 대한 위키피디아 문서를 연결한다. ERROR로 처리하지 않는다.

○ 예제

- (지명) 광진동
- <광진동: LCP_COUNTY>에 있는 마카롱 맛집 ‘날 웃게 해’
- (개체명) 광진동: LCP_COUNTY → (위키피디아) 광장동

개체명	위키피디아
광진동 LCP_COUNTY	<p>광장동</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>광장동(廣壯洞)은 서울특별시 광진구의 행정동, 법정동이다. 북쪽으로는 경기도 구리시, 동쪽으로는 한강, 남쪽으로는 구의3동, 서쪽으로는 구의2동과 접하고 있다.</p>

- 대상 문서에서 ‘광진동’이라는 언어 표현이 사용되었으나, ‘광진동’이라는 이름의 동은 존재하지 않는다. ‘날 웃게 해’라는 상점은 ‘광장동’에 위치하고 있음을 검색을 통해 확인할 수 있다. ‘광진동’은 ‘광장동’의 오타임이 명확하므로, 개체명 ‘광진동’에 위키피디아 ‘광장동’ 문서를 연결한다.
- 어떠한 언어 표현이 오자나 오타임이 명확한데, 해당 언어 표현도 단어로서 존재하며 위키피디아 문서도 있는 경우에는 대상 문서의 맥락에 맞는 위키피디아 문서를 연결한다. 즉, 언어 표현이 일치하는 위키피디아 문서를 연결하는 것이 아니라, 의미가 일치하는 위키피디아 문서를 연결한다.

○ 예제

- (지명) 상수동
- 서울숲 가야 해서 <상수동: LCP_COUNTY>으로 이동.
- (개체명) 상수동: LCP_COUNTY → (위키피디아) 성수동

개체명	위키피디아
상수동 LCP_COUNTY	<p>성수동</p> <p>위키백과, 우리 모두의 백과사전.</p> <p>성수동(聖水洞)은 서울특별시 성동구의 동으로, 법정동은 성수동1가, 성수동2가이고, 행정동은 성수1가제1동, 성수1가제2동, 성수2가제1동, 성수2가제3동으로 나뉘어 있다.</p>

- 대상 문서에서 ‘상수동’이라는 언어 표현이 사용되었으나, 맥락을 통해 ‘상수동’은 ‘성수동’의 오타임을 알 수 있다. ‘상수동’이라는 지명도 존재하지만, 대상 문서에서는 ‘상수동’이 아니라 ‘성수동’을 의미하므로 개체명 ‘상수동’에 위키피디아 ‘상수동’ 문서를 연결하는 것이 아니라 ‘성수동’ 문서를 연결한다.

(5) 개체명의 대상, 의미와 맞지 않는 문서가 자동으로 표시될 때의 처리

- 작업 도구에서 태깅된 개체명을 선택하면 해당 언어 표현을 제목으로 갖는 위키피디아 문서의 목록이 자동으로 보여지는데, 해당 개체명이 지시하는 대상과 도구에서 보이는 위키피디아 문서의 대상이 일치하지 않을 수 있다. 따라서 기계적으로 도구에서 보이는 위키피디아 문서를 연결해서는 안 되고, 검색을 통해 적합성을 판단한 후, 맞는 문서를 연결해야 한다.

○ 예제

- (지명) 홍대
- 주말에 <홍대: LC_OTHERS>에서 놀았어.
- (개체명) 홍대: LC_OTHERS / (작업 도구) '홍익대학교' 문서 제공 → (직접 검색) '홍대 (지명)' 문서 연결

개체명	위키피디아
홍대 LC_OTHERS	<p>홍대 (지명) 주A 9개 언어 ▾</p> <p>위키백과: 우리 모두의 백과사전.</p> <p> 홍대입구는 경기도 고양시에서 이 지역에 위치한 서울 지하철 2호선 수도권 전철 경의중앙선 및 인천국제공항철도의 환승역에 대해서는 홍대입구역 문서를 참조하십시오.</p> <p>홍대(弘大)는 서울특별시 마포구 서울 지하철 2호선 홍대입구역과 홍익대학교 캠퍼스를 잇는 지역 및 이 일대에 있는 문화가를 말한다. 홍대는 홍익대학교의 한국어 약칭으로 이 지역의 정식 지명이 아닌 통칭 명칭이며, 홍익대학교와 구분하기 위하여 홍대입구, 홍대거리라고도 부른다.</p> 

참고문헌

- 김나현(2021), 고유명의 다의성에 대하여, 『한국어 의미학』 74, 한국어 의미학회, 59-81.
- 김동환(2019), 『환유와 인지: 인지언어학적 접근법』, 역락.
- 민진우, 나승훈, 김현호, 김선훈, 강인호(2021) LUKE를 이용한 한국어 자연어 처리: 개체명 인식, 개체 연결, 『한국정보과학회 학술발표논문집』, 397-399.
- 박동호(2000), 대상부류 개념의 응용언어학적 활용 방안, 『프랑스어문교육』 10, 한국프랑스어 문교육학회, 35-52.
- 박재연(2008), 고유명의 변별적 속성에 대하여, 『한국어 의미학』 27, 한국어 의미학회, 103-126.
- 박재연(2012), 고유명 의미의 지시와 내포, 『한국어 의미학』 37, 한국어 의미학회, 183-211.
- 박재연(2014), 한국어 환유 표현의 체계적 분류 방법, 『한국어 의미학』 45, 한국어 의미학회, 1-30.
- 서반석(2016), ‘인명고유명사구’와 ‘지칭·호칭명사구’ 결합의 통사적 특성 연구, 『한국어학』 71, 한국어학회, 81-105.
- 유현조, 송영숙, 김민수, 윤기현, 정유남(2021) 딥러닝 기반 한국어 개체명 인식의 평가와 오류 분석 연구, 『언어』 46(3), 803-828.
- 윤평현(2013), 『국어 의미론 강의』, 역락.
- 이동혁(2007), 의미 범주 체계의 구축과 사전에서의 활용, 『한국어 의미학』 24, 한국어 의미학회, 51-82.
- 이동혁·이종열(2007), 고유명사의 비유적 의미 양상에 대한 연구, 『한국어 의미학』 23, 한국어 의미학회, 137-158.
- 이문성(1992), 고유명의 의미와 지칭, 『인문논총』 22, 전북대학교 인문학연구소, 227-246.
- 이성현(2005), 전자사전 구축과 의미부류 - 세종 명사 의미부류 체계의 예, 『한국어사전학』 5, 한국사전학회, 103-138.
- 이성현(2007), 세종 전자 사전의 어휘 의미 부류 체계, 『새국어생활』 17, 새국어생활, 51-82.
- 이영훈, 나승훈(2018) 위키피디아 링크 데이터를 이용한 한국어 뉴럴 멘션 탐지 및 개체명 연결, 『한국정보과학회 학술발표논문집』, 596-598.
- 이은경(2016), 정신 공간 이론과 고유명사, 『한국어 의미학』 51, 한국어 의미학회, 1-20.
- 이호경, 안재현, 윤정민, 배경만, 고영중(2017) 위키피디아 기반의 효과적인 개체 링크를 위한 NIL 개체 인식과 개체 연결 중의성 해소 방법, 『정보과학회논문지』 44(8), 813-821.
- 이흥식(2016), 고유명사의 의미와 지시에 대하여, 『한국학연구』 40, 인하대학교 한국학연구소,

307-334.

임지룡(2006), 환유 표현의 의미특성, 『인문논총』 55, 인문학연구원, 265~300.

임홍빈·한재영(1993), 국어 어휘의 분류 목록에 대한 연구, 국립국어연구원.

장석배·강민지·김진웅(2020), 환유 연구를 위한 말뭉치의 구축과 주석: 신문 말뭉치에 나타난 도시명의 환유 주석을 중심으로, 『언어사실과 관점』 50, 연세대학교 언어정보연구원, 75~98.

전상호(2021), 개체명으로서의 인명의 특성과 분류 체계에 대하여, 『우리말글』 88, 우리말글 학회, 31-59.

차정원(2019), 개체명 분석 말뭉치 구축, 국립국어원.

차정원(2020), 2020년 개체명 말뭉치 연구 분석, 국립국어원.

차준경(2003), 고유명사의 의미와 중의성 해소, 『한국어학』 20, 한국어학회, 265-286.

최경봉(1996), 명사의 의미 분류에 대하여, 『한국어학』 4, 한국어학회, 11-45.

최경봉(2001), 지식기반 구축을 위한 어휘의 의미 부류, 『담화와인지』 8, 담화와인지언어학회, 275-303.

허정(2020), Entity Linking 가이드라인, 한국전자통신연구소.

<Abstract>

Corpus Research and Analysis 2021: Named Entity Tagging and Entity Linking

In the era of the fourth industrial revolution and the contactless environment, Big Data and Artificial Intelligence Technology have been in the limelight. Large-scale machine learning data have been increasingly in demand, and National Institute of Korean Language has been building various corpora for both companies and university laboratories. In this project, we focus on constructing a named entity corpus and linked data. In addition, we suggest the criteria and direction of the next project by clearly establishing guidelines for the construction.

‘Corpus Research and Analysis 2021: Named Entity Tagging and Entity Linking’ comprises two tasks: named entity analysis and entity linking. The named entity analysis consists of the following subtasks: 1) annotating named entity in the written and colloquial raw corpus of 3 million words, 2) revising the annotation guidelines for named entities, and 3) expanding the previously constructed corpus with 15 named entity tags to the one with 150 tags.

The entity linking task consists of the following subtasks. 1) developing the methodology for entity linking, 2) making the guidelines, and 3) building the entity linked data of 11 million words in total.

This project is conducted by the Data.ly consortium: Data.ly a lead partner and Hallym University a consortium partner. Data.ly has responsibility for the overall management of the corpus construction process and system, tool development, and annotator management and education. Hallym University revised and developed the named entity and entity linking annotation guidelines.

Keywords: Named Entity, Entity Linking, Guidelines, Corpus Construction
Project Director: Cha Jeong-won (Data.ly)

<기획·연구>

국립국어원 이승재 언어정보과장

국립국어원 서셋별 학예연구사

국립국어원 오은비 연구원

<사업 참여자>

사업 책임자 차정원(데이터리 대표)

사업 참여자 신서인(한림대학교 교수)

오진영(데이터리 이사)

이한범(한림대학교 시간강사)

이수현(한림대학교 시간강사)

박진현(한림대학교 시간강사)

전상호(한림대학교 시간강사)

김나현(한림대학교 박사수료)

김희인(한림대학교 연구원)

최양선(한림대학교 석사)

남장현(한림대학교 석사과정)

홍진기(한림대학교 석사과정)

조아임(경희대학교 박사수료)

오수진(경희대학교 박사수료)

임윤정(명지대학교 시간강사)

황지영(경희대학교 석사)

김희성(경희대학교 석사)

김지영(경희대학교 석사과정)

임고은(경희대학교 석사과정)

이상금(한림대학교 시간강사)

박은실(데이터리 전임연구원)

배진영(데이터리 전임연구원)

최윤진(데이터리 전임연구원)

서강원(데이터리 전임연구원) 외 32인

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2021년 12월 17일

발행일: 2021년 12월 17일

인 쇄: 서경기획

※ 이 책은 국립국어원의 용역비로 수행한 ‘2021년 개체명 분석 및 개체 연결 말뭉치 연구 분석’ 사업의 결과물을 발간한 것입니다.