

국립국어원 2019-01-14

발 간 등 록 번 호
11-1371028-000773-01

개체명 분석 말뭉치 구축

사업 책임자 차 정 원

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 ‘개체명 분석 말뭉치 구축’에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업기간: 2019 년 7 월 ~ 2019 년 11 월

2019 년 11 월 22 일

사업 책임자: 차 정 원 (데이터리)

사업 수행 기관: (주)데이터리, 창원대학교

사업 책임자: 차정원

공동 연구원: 정경재

연구원: 오진영, 박은실, 배진영, 최윤진

<사업 수행자> 데이터리, 창원대학교 산학협력단

사업 책임자	차정원 (데이터리 대표, 책임연구원)
사업 참여자	정경재 (창원대학교 국어국문학과 조교수, 공동연구원)
	오진영 (데이터리 이사, 참여연구원)
	박은실 (창원대학교 산업기술연구원 전임연구원, 참여연구원)
	배진영 (데이터리 전임연구원, 참여연구원)
	최윤진 (데이터리 전임연구원, 참여연구원)

개체명 분석 말뭉치 구축

본 사업은 4차 산업혁명에 대비하여 인공지능 언어처리 기술 개발 및 국어 연구의 기반 자료로 활용할 수 있는 대규모 고품질 개체명 분석 말뭉치를 구축하기 위한 사업이다. 개체명 분석 말뭉치 구축에 적합한 프로그램과 전문가의 엄밀한 검증 및 정제를 통하여 대규모 고품질 말뭉치를 구축하는 데 목표를 두었다. 또한, 개체명 구축에 사용되는 한국전자통신연구원(ETRI, 이하 ETRI)의 ‘세부분류 개체명 가이드라인 2018’ 지침을 숙지하고 작업을 진행하는 도중 발견되는 지침의 수정사항 및 보완점 등을 국립국어원에 전달하여 기존의 지침을 보완한, 좀 더 향상된 지침을 완성하는 데 일조하고자 하였다.

사업 기간은 약 4개월이며 작업 대상은 국립국어원에서 전달한 원시 말뭉치 300만 어절이다. 사업 기관인 데이터리와 공동사업기관인 창원대학교는 국립국어원으로부터 문어 원시 말뭉치 200만 어절과 구어 원시 말뭉치 100만 어절을 전달받아 작업을 진행하였다. 작업은 5% 이하의 오류율을 목표로 하였다.

효율적인 방식으로 성공적인 결과를 끌어내기 위하여 본 사업은 다음과 같이 진행되었다.

첫째, 데이터리에서 확보하고 있는 다양한 프로그램을 활용하였다. 일관성 검사를 기반으로 하는 레이블링·검증 도구와 특허 인증 기술인 Rough set을 이용한 오류 판별 방법, RDR을 기반으로 하는 오류 자동 수정 방법 등 자동화 시스템을 통하여 작업을 수행하였다. ETRI 지침에서 분류한 15가지 개체명 범주를 자동화 시스템에 적용한 후, 개체명 말뭉치의 구축·정제·오류검출 작업을 진행하였다. 자동화 시스템의 작업 과정은 에러 검출부터 수정까지 전문가의 관리하에 일괄적으로 이루어지므로 전문성과 정확성을 모두 얻을 수 있는 방식이다. 이 작업을 통해 산출된 결과물을 전문가가 다시 한번 샘플링을 통하여 검토하고 수정하였다. 또한, 작업자에 의해 구축된 모든 결과물은 자동화 시스템의 성능 향상을 위한 자료로 다시 이용되었다. 작업 중 발생한 오류-수정 쌍을 이용하여 오류 수정 규칙을 자동으로 생성시키고 시스템 내 전체 말뭉치에 적용하여 일괄 수정 작업을 거쳐 최종적으로 말뭉치의 성능을 향상시키는 방식을 사용하였다.

둘째, 개체명 말뭉치를 구축하고 수정하는 작업자들을 고용하여 동일한 자료를 세 차례 검토하도록 하였다. 모든 작업자는 본 작업을 시작하기 전 전문가에 의한 지침 교

육을 받았으며, 샘플 말뭉치를 통해 지침을 숙지하는 단계를 거쳤다. 개체명 말뭉치를 구축하는 도구로는 데이터리에서 보유하고 있는 프로그램 ‘Kane’을 사용함으로써 작업의 효율성을 높이고 단순 실수를 낮췄다. 고용된 작업자들은 자동화 시스템을 통해 전처리된 말뭉치를 일일이 검토하면서 수정하였다. 또한, 중도에 발생하는 작업자들의 질의 사항에 대해 전문가가 답변하고 이를 공유함으로써 반복적으로 지침에 대해 재교육하였다. 그와 함께 국어학 전문가를 중심으로 한 품질보증팀을 별도로 꾸려 1차 작업자의 결과물을 다시 검토하고 수정하도록 하였다. 이를 통해 오류율을 줄이고 말뭉치의 완성도를 높였으며, 검토 과정에서 발생한 전문가의 의견은 지침에 반영될 수 있도록 따로 정리하여 국립국어원에 전달하였다.

마지막으로 사업팀의 모든 결과물, 즉 개체명 분석 말뭉치 및 작업 도중 발생한 질의 사항과 지침 수정 건의 사항을 국립국어원에 전달하고 국립국어원이 이를 검토하도록 하였다. 결과물에 대한 오류와 수정사항을 국립국어원에서 전달받은 후 그에 대한 수정을 진행하며, 양자간 논의가 필요한 사안에 대해서는 의견을 조정한 후 이를 지침에 반영하여 지침의 완성도를 높였다.

본 사업을 통하여 완성된 작업물은 민간에서 자유롭게 활용할 수 있도록 공유재로 사용될 것이며, 이로써 인공지능 언어처리 기술 개발을 위한 기반이 될 것이다. 또한, 정보화 시대에 필수적인 국어 및 국어문화 연구를 위한 기초 자료 구축의 의의를 지니며, 사업 과정에서 발생한 여러 논의 사항들은 국어 정책 수립에도 도움이 되리라 기대한다.

말뭉치 구축 사업은 많은 이들이 주목하고 있으며 현재보다 미래에서 그 중요성이 더욱 부각될 것이다. 본 사업의 수행 기관인 데이터리와 창원대학교는 국어 연구 활동과 관련 산업의 발전에 도움이 되고자 국립국어원이 진행한 ‘개체명 분석 말뭉치 구축’ 사업에 참여하였다.

주요어: 자동화 시스템, 개체명, 말뭉치, 검증, 지침

차 례

제1장 사업 개요

1. 사업 목표 및 기대 효과	3
2. 사업 수행 방법 개관	4
3. 중간 및 최종 산출물	5
4. 향후 논의	6

제2장 개체명 분석 말뭉치 구축 과정 및 방법

1. 사업 수행 방식	9
2. 개체명 분석 말뭉치 구축 도구	11
3. 말뭉치 구축 인력	15

제3장 개체명 분석 말뭉치 구축 지침

1. 대원칙 보완	25
2. 전체 부류에 적용되는 일부 지침 추가	27
3. 부류별 지침 수정	28

제4장 결론

1. 사업 요약	33
2. 남은 문제	34

참고문헌

부록. 개체명 분석 말뭉치 구축 지침 2019

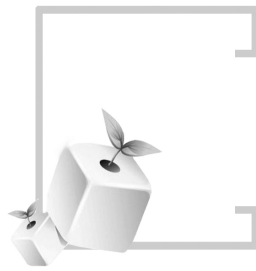
표/그림 차례

표 목차

- <표 1>	23
- <표 2>	24

그림 목차

- <그림 1>	3
- <그림 2>	9
- <그림 3>	12
- <그림 4>	13
- <그림 5>	13
- <그림 6>	14
- <그림 7>	15
- <그림 8>	17
- <그림 9>	18
- <그림 10>	18



제 1 장

사업 개요



제1장 사업 개요

1 사업 목표 및 기대 효과

본 사업은 다음과 같은 목표를 기반으로 하여 진행되었다.

첫째, 국립국어원에서 제공한 원시 말뭉치 300만 어절을 대상으로 개체명 분석 말뭉치를 구축하는 것을 목표로 한다. 오류율은 5% 이내로 설정하여 고품질의 말뭉치를 구축한다.

둘째, 개체명 분석 말뭉치 구축 작업을 위한 지침을 정리한다. 기존에 ETRI에서 작성 및 사용하던 지침을 수정·보완하여, 향후 관련 사업에서도 본 지침이 사용될 수 있도록 완성도 있는 지침을 마련한다.

본 사업팀은 사업 진행 과정 전반을 위 목표에 맞게 설정하여 사업 결과물이 목표에 부합하도록 노력하였다. 사업의 목표대로 원시 말뭉치 300만 어절의 구축과 지침의 수정 및 보완은 사업 종료와 함께 성공적으로 완료되었으며 본 사업의 결과물이 가져올 기대효과는 다음과 같다.



<그림 1> 사업 기대효과

본 사업은 미래 전망적인 인공지능 관련 사업과 4차 산업의 발전에 기반이 될 대용량 말뭉치 구축 자료를 생성한다. 사업을 통해 생성된 작업물은 민간에 공유되며 현재 주목받고 있는 기술인 언어 인공지능 산업과 빅데이터 산업 등 관련 사업에 중요한 토대로 쓰일 것이다. 언어 인공지능 산업과 빅데이터 산업의

생명은 풍부한 데이터와 그 데이터에 대한 정확도이기 때문에 본 사업의 결과물은 산업의 발전에 아주 큰 역할을 할 것이 분명하다.

또한, 그 중요성이 날이 갈수록 더해지고 있는 국어 및 국어문화 연구와 국어 정책 수립의 기초 자료로 활용되어 국어 연구 환경에 도움을 줄 것이다. 관련 자료가 부족한 현시점에 대용량 개체명 구축 자료는 다음 단계로의 발전에 이바지할 수 있는 디딤돌이 될 것이다.

사업의 진행 과정에서 말뭉치 구축 지침 계속해서 수정과 보완을 거쳐 전보다 완성도 높은 지침으로 거듭날 것이다. 말뭉치 구축 지침은 추후 관련 사업에서도 장기적으로 사용되어야 하는 중요한 자료이다. 사업을 통해 만들어진 완성도가 높은 지침은 향후 관련 연구 진행에서도 본보기로 사용되어 사업이나 연구가 성공적으로 마무리될 수 있도록 도움을 줄 것이다.

2 사업 수행 방법 개관

본 사업의 참여 기관은 데이터리(Data.ly)와 창원대학교 두 기관이다. 고품질의 말뭉치를 구축하기 위해, 그리고 말뭉치 구축과 함께 지침 수정·보완 작업을 병행하기 위해 두 참여 기관은 각 팀에서 전문적으로 맡을 수 있는 과업을 중심으로 분업하여 효율적으로 사업을 진행하였다.

데이터리는 사업의 전반을 총괄하는 한편, ‘말뭉치 구축팀’으로서의 역할을 수행하였다. 데이터리의 전문가가 개발한 자동화 시스템을 이용하여 개체명 말뭉치를 구축하고, 말뭉치의 오류를 일일이 보면서 수정할 작업자를 모집하여 1차 수정 말뭉치 구축을 진행하였다. 수작업으로 진행되는 말뭉치 수정작업에 필요한 프로그램 ‘Kane’을 제공하여 작업에 효율성을 높였다. 작업자에 의해 수정이 완료된 데이터는 다시 자동화 시스템을 이용하여 오류를 검출해 낸 뒤 문서를 정제하여 품질보증팀에게 전달하였다.

창원대학교는 국어학적 지식을 지닌 전문가가 중심이 되어 ‘품질 보증팀’의 역할을 수행하였다. 말뭉치 구축팀에서 전달받은 문서를 검토하고 오류를 수정하여 2차 수정 말뭉치를 구축한 후 이를 다시 말뭉치 구축팀에 전달하였다. 품질 보증팀에서 검출해 낸 오류 사항들은 말뭉치 구축팀의 자동화 시스템에 적용되어 프로그램의 성능을 향상시키는 데 기여하였다. 또한, 지침에서 발견되는 오

류나 의문 사항들을 국어원에 전달하며, 지침을 수정·보완하는 작업을 진행하였다.

3 중간 및 최종 산출물

본 사업의 중간 및 최종 결과물은 다음과 같다.

□ 중간 산출물

납품명	내용
개체명 분석 말뭉치 샘플	사전 검토를 위한 샘플링 문서 10건
개체명 분석 말뭉치 샘플	시범 검증을 위한 2만 어절 (73개 문서)
개체명 분석 말뭉치 (1차)	구문 분석 말뭉치 문어 88만 어절 (30%)

□ 최종 산출물

납품명	내용	매체
개체명 분석 말뭉치 (최종)	구문 분석 말뭉치 문어·구어 300만 어절 (100%)	외장 SSD 2대
개체명 분석 말뭉치 지침	구문 분석 말뭉치 구축 시 문제점이 수정·보완된 지침	
최종 보고서	사업 수행 과정 및 결과 요약 보고서	인쇄본 30부

□ 개체명 분석 말뭉치 구축 지침

본 사업은 기존에 공개되었던 ETRI 지침을 출발점으로 삼아 시작되었으며, 말뭉치 구축 작업과 함께 지침 수정과 말뭉치 재수정 작업이 병행되며 진행되었다. 기존의 ETRI 지침은 ‘개체명에 어떠한 태그를 부착할 것인가’ 하는 문제에 집중되어 있었으며, ‘무엇이 개체명인가’, ‘어떠한 단위로 태깅할 것인가’ 하는 문제에는 소홀했다. 또한 구어 말뭉치를 고려한 지침도 없었다. 따라서 이러한 미비점을 크게 보완할 필요가 있었으며, 지침 전체의 일관성을 잡고 소소한 오류를 수정하는 작업도 함께 진행하였다. 사례 중심의 설명을 일반화된 기술로 바꾸는 작업도 필요하였다.

본 사업팀에서는 작업 중에 마주하는 기존 지침의 한계를 정리하고 안을 제시

하여 사업 주최 기관인 국립국어원에 전달하였으며, 수차례 논의를 통해 의견을 조율한 결과를 지침에 반영하였다. 지침의 마지막 개체명 성능 평가 부분도 수정하여 최종 확정된, 수정·보완 지침을 국립국어원에 납품하였다.

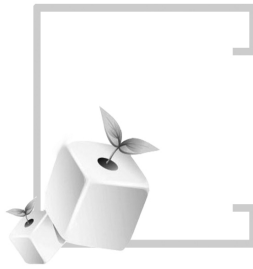
4 향후 논의

앞의 내용을 통해 ‘개체명 분석 말뭉치 구축’ 사업의 전반적인 내용에 관해 기술하였다. 이후 본 보고서의 내용은 다음과 같이 진행된다.

제2장 ‘개체명 분석 말뭉치 구축 과정 및 방법’에서는 사업을 이행하기 위하여 본 기관에서 자체적으로 수립한 방식과 일정 및 계획에 대하여 논하고, 작업에 주로 사용되었던 프로그램에 대하여 간략히 설명하고자 한다. 그 후에는 작업에 참여하였던 본 기관의 조직도 및 구성원을 소개하며 사업에 참여한 작업자들이 구체적으로 어떤 일을 하였고, 작업의 과정은 어떻게 관리하였는지 기술한다.

제3장 ‘개체명 분석 말뭉치 구축 지침’에서는 지침의 내용을 검토하며 변경된 부분을 중심으로 수정 및 보완이 이루어진 부분에 대해 상세히 설명한다. 전체적으로 지침의 대원칙 보완, 그리고 추가 및 신설된 부분에 대한 설명, 세세한 항목별로 수정된 부분으로 나누어 기술한다.

마지막 장에서는 앞서 기술하였던 사업의 내용을 요약하고 마지막으로 남은 문제들에 대하여 기술하며 본 보고서를 마무리할 것이다.



제 2 장

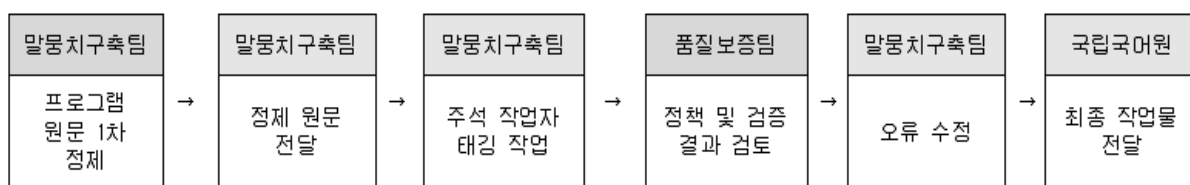
개체명 분석 말뭉치 구축 과정 및 방법



제2장 개체명 분석 말뭉치 구축 과정 및 방법

1 사업 수행 방식

□ 진행 과정



<그림 2> 사업 진행 과정

성공적인 사업 진행을 위하여 말뭉치 구축팀과 품질 보증팀의 역할에 맞게 과정별로 업무를 분담하였으며 작업의 단계는 크게 4단계로 구분하였다. 과정은 다음과 같이 진행되었다.

첫 번째는 작업단계이다. 이 단계에서는 사업 기관인 데이터리(Data.ly)-창원대학교와 사업 주최 기관인 국립국어원 간의 미팅이 진행되며 말뭉치 지침을 분석하고 이에 대한 논의 및 협의를 진행한다. 또한, 말뭉치 구축팀에서는 말뭉치 구축을 위한 전처리, 구축, 평가 검증 시스템을 다시 한번 정비하고 바로 적용할 수 있도록 준비하여 원문의 1차 정제가 원활히 이루어질 수 있도록 한다.

다음은 구축단계이다. 구축단계에서는 말뭉치구축팀에서 보유하고 있는 자료를 바탕으로 학습을 끝낸 시스템을 이용하여 초기 말뭉치를 구축한다. 말뭉치구축팀은 보유하고 있는 자동화 시스템을 이용하여 문서 원문을 1차 정제한다. 이 과정에서 프로그램이 이전에 학습하였던 개체명 분석 자료를 바탕으로 원문에 자동으로 개체명이 부착된다. 이 과정을 거친 1차 정제 원문을 개체명 부착 작업자들에게 전달한다. 지침을 숙지한 작업자들은 프로그램이 잘못 부착한 개체명과 부착하지 않은 부분을 ‘kane’ 프로그램을 이용하여 수정하고 부착한다. 이후 작업자들의 부착 작업 결과물을 자동 오류 탐지 시스템에 적용시켜 오류를 검출해내고 이를 수정하는 피드백 작업을 실행한다. 위 과정은 계속 반복되며 이를 통해 지침의 재정리 및 정비에 대한 발판 또한 마련한다.

검증단계에서는 지침 숙지를 끝낸 전문가가 작업한 샘플링 평가를 통한 다면 평가를 진행한다. 작업자의 손을 거친 2차 정제 문서는 마지막 검토를 위해 품

질 보증팀으로 전달되고, 국어학적 지식을 지닌 전문가가 팀원들과 함께 꼼꼼히 검토한 후, 오류를 검출해낸다. 오류가 많을 경우 검증 결과를 말뭉치 구축팀에 전달하고 작업자들이 이를 수정하게 하며, 오류 사항들을 자동화 시스템에 적용하여 시스템의 성능을 높인다. 오류에 대해 새로 학습한 시스템은 문서에 재적용되며 작업의 정확성과 일관성을 높이게 된다. 전문가의 구축 작업과 오류검출 작업이라는 두 가지 평가를 통해서 생성된 새로운 규칙은 자동화 시스템에 새로 학습되며 전체 말뭉치에 적용된다. 위 과정은 개체명 구축과 병행하여 시행한다.

마지막으로 종료 단계에서는 작업한 말뭉치 구축 결과를 정리하고 이를 분석한 통계 정보를 도출해낸다. 그리고 위 3단계를 거쳐 완성된 작업물을 국립국어원에 전달한다. 전달된 작업물은 국립국어원 자체 검증팀에서 검증 작업이 이루어지며 검증이 끝난 뒤 그에 대한 결과는 말뭉치 구축팀과 품질 보증팀에게 전달된다. 마지막 검증 과정에서 발생한 오류나 지침 변경으로 인해 수정되어야 하는 부분들을 최종 보완하고 사업에 대한 최종 보고를 마치면 ‘개체명 분석 말뭉치 구축’ 사업이 종료된다.

□ 연구 공정표

연구 공정표						
일련 번호	개발내용	추진 일정				
		7	8	9	10	11
1	<ul style="list-style-type: none"> ■ 말뭉치 지침 분석 ■ 말뭉치 검증 시스템 준비 					
2	<ul style="list-style-type: none"> ■ 말뭉치 구축 ■ 오류 탐지 ■ 오류 수정 					
3	<ul style="list-style-type: none"> ■ 다면평가 ■ 통계적 평가 					
4	■ 말뭉치 구축 마무리					
4	■ [산출물] 개체명 분석 지침					
5	■ [산출물] 최종 보고서					

2 개체명 분석 말뭉치 구축 도구

□ 사업 추진 전략 및 지원체계

본 사업 목표를 위하여 ‘품질 보증팀(이하 보증팀)’과 ‘말뭉치 구축팀(이하 구축팀)’의 유기적인 지원이 필요하다. 구축팀은 지침에 맞게 말뭉치를 구축하는 것이 목표이고 보증팀은 지침의 보완, 수정, 구축 및 구축팀 결과에서 수정이 필요한 항목과 누락된 항목 등을 전달하여 구축팀과 자동화 시스템에 반영하는 역할이다.

보증팀과 구축팀 사이에 자동화 시스템은 고품질 말뭉치 구축을 위한 도구를 제공하고 먼저 학습된 전문가의 지식을 통해 구축팀의 작업을 보조하며, 보증팀의 피드백을 반영하는 역할을 한다.

□ 점진적 개선 프로세스

본 사업에서 개체명 분석 말뭉치 구축을 위하여 점진적으로 오류를 줄여나가며 구축하는 프로세스를 설계하여 적용한다. 기존 대용량 분석 말뭉치를 구축시 어려움은 말뭉치 구축에서 작업자가 지침을 이해하고, 원시 말뭉치를 할당받아 분석 말뭉치를 구축하며 고품질을 위하여 수정작업을 진행한다. 이와 같은 진행방식은 구축팀과 보증팀 역할을 동시에 하게 되어 구축에 필요한 시간 및 비용이 많아지게 된다.

본 사업에서는 이러한 문제를 보완하고자 전문가가 구축한 지식을 학습한 자동화 시스템으로 원시 말뭉치를 분석하고, 초기 구축 작업을 자동으로 수행한다. 자동화 시스템의 초기작업은 각 문장의 자질을 추출하고 자질에 일치하는 전문가 지식을 검색한다. 자질은 관찰이 필요한 문장 내 특징적인 정보를 선택하는 것으로 예를 들면 문장의 형태소 분석 결과 중 체언을 선택하거나 숫자 정보들이 포함되어 있는지 등이 해당한다. 사람과 달리 자동화 시스템은 각 문장에 대하여 수백에서 수천 가지의 자질을 선택하고 비교하는 작업을 함으로써 빠른 속도로 초기작업을 수행한다.

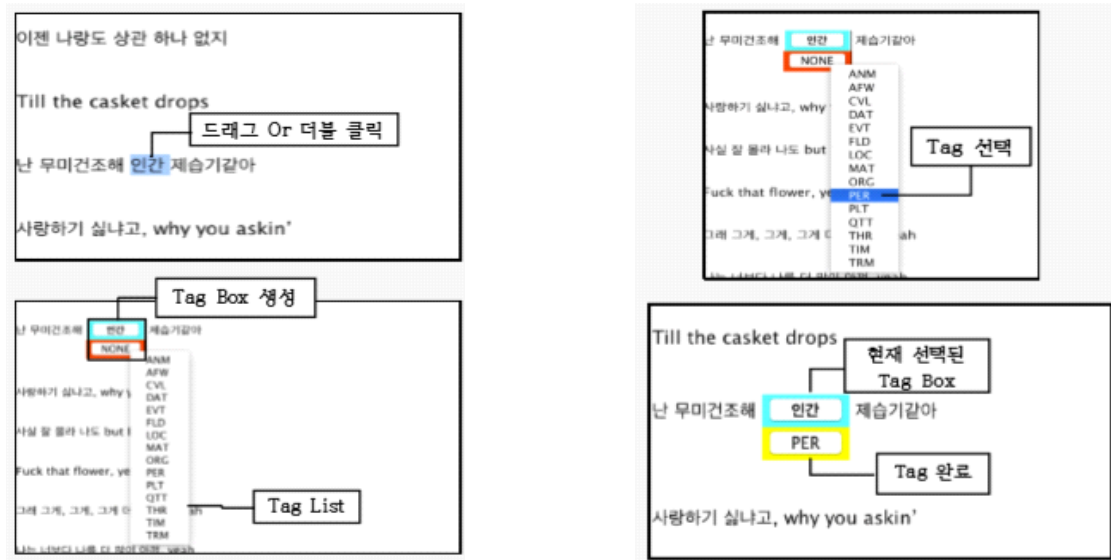
또한, 자동화 시스템은 초기작업 결과에 대하여 각 문장에 대한 예상 정확도를 계산하여 정확도가 높지 않은 문장만 추출하는 기능(샘플링)을 가지고 있다. 예상 정확도가 높지 않은 문장은 문맥에 따라 변화가 가능한 개체명이 포함된 경

우이다. 이러한 문장을 구축팀에 전달하여 작업함으로써 구축팀의 작업량을 줄이는 효과를 준다. 그리고 구축팀이 작업한 말뭉치에 대하여 보증팀이 수정하면 둘 작업 결과를 비교하여 자동화 시스템이 가지고 있는 지식에 반영하게 된다.

이러한 프로세스(초기 구축 작업 - 샘플링 - 구축팀 작업 - 보증팀 검증)를 여러 번 반복하는 과정을 점진적인 개선 프로세스로 정의하고 이 과정을 통하여 자동화 시스템의 분석 성능과 개체명 구축 말뭉치의 품질도 높일 수 있다.

□ 말뭉치 구축 및 수정 작업방법 - kane 툴

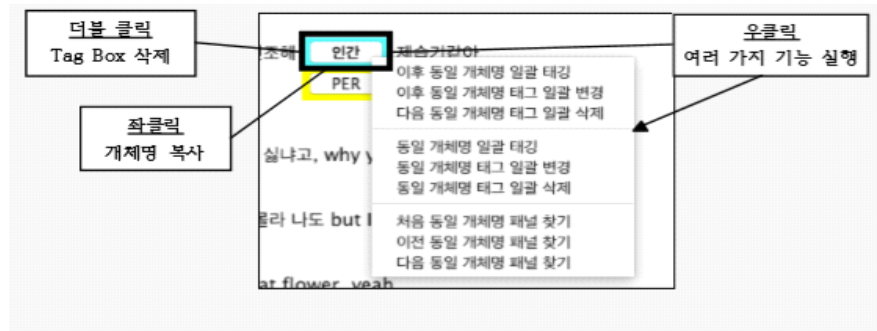
말뭉치 구축 작업 시 시간 단축과 정확성을 위하여 개체명 부착 및 수정 프로그램 ‘kane’을 사용한다. ‘kane’은 말뭉치 구축 작업자가 선택한 개체명에 지침 속 15가지의 범주 중 적절한 하나의 태그(tag)를 말뭉치에 부착하는 데이터리 자체 제작 프로그램이다. 자동화된 시스템의 결과물 중 자동화 시스템만으로 모호성을 해소할 수 없는 문장들로 선별된 것들을 분배하고, 구축 작업자는 kane 툴을 이용하여 작업한다.



<그림 3> 구축 도구 kane 사용 방법

사용 방법은 간단하다. 작업이 필요한 문서를 툴 내에서 연 다음, 문서에 포함된 개체명을 선택하여 개체명에 부합하는 태그를 선택하는 것이다. 개체명을 선택하는 방법은 문장에서 개체명에 해당하는 구간을 드래그하거나, 어절이 개체명이면 더블클릭하여 선택하는 방법이 있다. 그러면 선택된 구간에는 두 개의

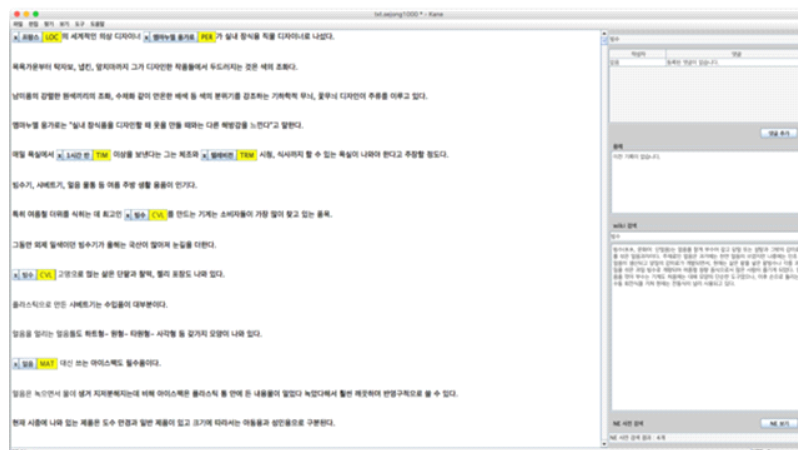
박스(box)가 표시된다. 하나의 박스는 선택된 개체명 어휘를 나타내고, 나머지는 개체명 태그 15개 중 하나를 선택할 수 있다. 개체명이 선택된 부분은 문장 내 개체명 작업 진행이 직관적으로 확인할 수 있도록 비 개체명과 구분하여 표시된다.



<그림 4> 구축 도구 kane 사용 방법 2

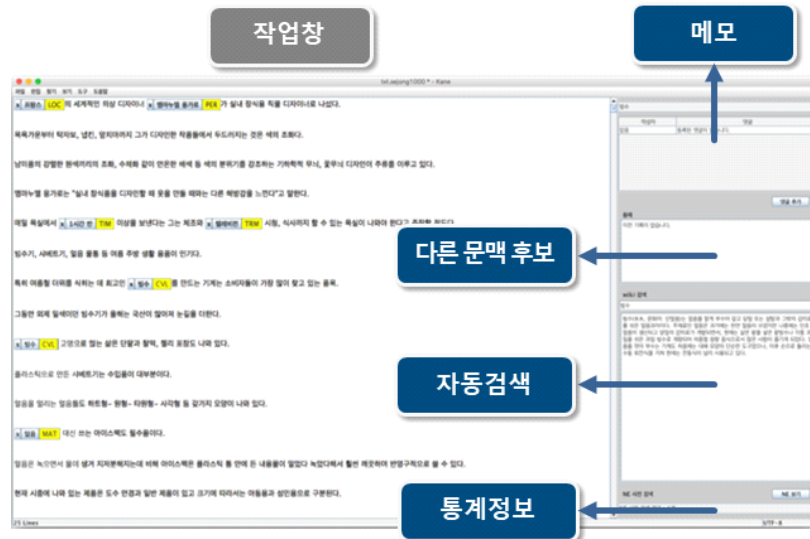
개체명 구간이 잘못 선택된 경우 더블 클릭을 하여 제거할 수 있다. 또한, 현재 문서 내 같은 개체명이 다수 등장하거나, 문서 내 동일 개체명 작업 결과에 대하여 일괄 검색이나 변경, 또는 삭제가 필요한 경우 몇 번의 선택으로 기능 수행이 가능하다.

초기 원시 말뭉치는 xml 포맷으로 자동화 시스템 또는 작업자에게 원 포맷을 이용할 경우 오류 발생 가능성이 커지고 불편하다. 이러한 문제를 해결하기 위하여 자동화 시스템 및 kane 툴에서는 자체 설계한 포맷을 사용한다. 자체 설계한 포맷은 기본적인 문서의 형태이며 이 형태의 파일을 kane 툴을 통하면 개체명이 구분된 아래와 같은 화면이 보여진다.



<그림 5> kane에서 문서를 열었을 때 화면

구축 작업은 문서 하나의 작업이 아니라 다수의 문서와 다수의 구축 작업자의 일관성 있는 작업 결과물이 중요하다. 이를 위하여 Kane 툴에서는 다음과 같은 기능을 제공한다.



<그림 6> kane의 세부 기능

첫 번째는 메모 기능이다. 작업 창에서 개체명을 선택하면 오른쪽 최상단에 메모 기능이 활성화된다. 메모에는 구축 작업자가 개체명에 대하여 문의, 생각 등을 공유할 수 있다. 공유된 메모를 통하여 의견을 나눌 수 있으며 오류를 감소시키는 효과를 준다.

두 번째 기능은 문맥 정보 확인 기능이다. 개체명의 경우 문맥에 따라, 같은 개체명이지만 태그가 달라지는 경우가 있다. 각 문맥에 따른 개체명 태그와 예제를 공유하는 기능이다. kane 툴의 오른쪽 중간에 표시된다.

마지막은 자동 검색 기능이다. 개체명 작업 시 사전을 참조하여 의미, 정의를 확인하는 일이 많다. 개체명을 선택할 시 해당 개체명의 사전적 의미를 자동으로 검색하여 표기하여 편의성을 제공한다.

'kane'을 사용함으로써 작업자들은 화면 속 작업 창에서 마우스 클릭으로 개체명 작업을 수행하므로 입력 오류를 줄일 수 있다. 또한, 위에서 설명한 다양한 편의 기능이 포함되어 있으므로 부착 시간과 오류를 줄이는 데 효과적이다. 또 작업 창에서 개체명을 선택하는 경우 오른쪽에 메모, 문맥 정보, 사전 정보를 실시간으로 보여 줌으로써 개체명 부착 작업의 효율을 높인다. 그리고 개인적으로 작업이 이루어지지만 다른 작업자들의 결과물을 확인할 수 있는 점을 통해 공동

작업자들이 참조할 수 있게 되어있어 협업이 가능하다.

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

□ 말뚱치 정책 적합 여부 검토 도구

사업 진행 과정에서 국립국어원과 의견을 교류하며 지침상 수정이 이루어지는 경우가 발생한다. 그로 인해 사업의 정책이 변경될 경우 기존에 이미 작업된 결과물을 변경하는 작업이 필요하다. 작업을 위해 자동 시스템에 반영하여 기존 프로세스로 진행할 경우, 반영에 많은 시간이 소요된다. 또한, 구축 작업자의 혼란도가 높아져 오류 발생 가능성이 커진다. 따라서 말뚱치 정책 적합 여부 검토 작업을 진행한다.

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

이 작업은 변경된 정책 적합 여부뿐만이 아니라 최종 검토 시 검토 목록을 작성하여 반영 여부를 확인하고 최종 결과물의 품질을 높인다.

3 말뚱치 구축 인력

□ 사업 참여 기관



<그림 7> 사업 참여 기관

본 사업은 국가기관인 국립국어원이 주최하는 대규모의 국가사업으로 위의 네 기관이 참여하여 서로 끊임없는 소통과 협력이 이루어져야만 성공적인 결과를

이루어 낼 수 있다. 그리하여 실제 각 기관이 맡은 역할에 충실하고 서로 지속적인 교류를 이어갔으며, 이 결과 사업이 원활히 마무리될 수 있었다. 각 기관이 이행하였던 실제 과업 내용은 다음과 같다.

‘개체명 분석 말뭉치 구축’ 사업의 주최 기관인 국립국어원은 사업 이행 기관이 사업의 목적 및 목표를 뚜렷하게 인식할 수 있도록 사업의 전체적인 방향과 세부적인 내용을 명확히 설정한다. 전체적인 사업의 주관으로서 용역 기관이 제출한 최종 결과물을 판단하고, 전체적인 완성도를 자세히 검토한 후 사업의 마무리를 결정한다.

외부검증팀은 국립국어원 내부에 조직된 기관으로 용역 기관에서 제출한 개체명 구축 작업물에 대하여 검증을 진행한다. 검증은 자체 개발한 시스템을 이용하여 이루어지며 검증통계와 검증내용, 오류 내용과 오류율 등을 산출하여 용역 기관에 전달하여 수정작업 시 필요한 정보들을 제공한다.

데이터리(말뭉치 구축팀)는 사업 참여 기관으로서 맡게 된 ‘개체명 분석 말뭉치 구축’의 책임과 전체적인 일정 및 작업을 관리한다. 자동화 시스템과 인력을 이용한 개체명 구축 작업을 진행하며, 전체 총괄 기관으로서 결과물에 대한 품질관리에 책임이 있다.

창원대학교(품질 보증팀)는 작업한 문서의 수정과 검토를 통해 오류를 잡아내거나 지침의 문제 사항들을 잡아낸다. 이 작업물을 말뭉치 구축팀에 전달함으로써 결과물의 내부 검증을 진행하는 동시에 말뭉치구축팀이 보유하고 있는 도구의 성능을 향상시키는 데 일조하며, 국어학적 지식을 지닌 집단으로서 언어에 대한 지식을 제공한다.

□ 수행조직 및 업무 분담

사업 기관은 데이터리(Data.ly)이며, 공동 사업 기관으로 창원대학교 산학협력단이 참여한다. 하위 부서는 사업지원팀, 품질보증팀, 말뭉치구축팀, 기술지원팀으로 이루어져 있으며 부서의 역할은 다음 설명과 같다.

사업지원팀은 본 사업의 총괄을 맡는다. 데이터리의 대표이자 사업의 총괄을 맡은 차정원 대표가 팀장으로 있으며 사업의 시작부터 마무리까지의 진행 일정, 운영 방식, 공정 관리 등 사업 진행의 모든 절차를 관리·감독한다. 지침의 해석이나 수정에 대한 문제로 국립국어원과의 논의가 필요할 경우, 우선 본 기관의

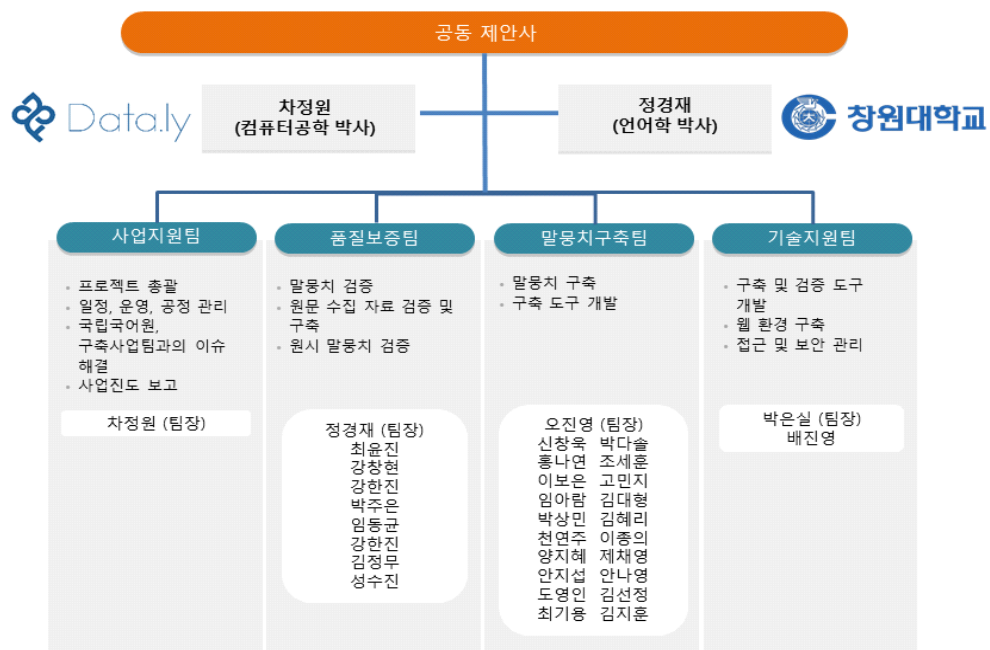
의견을 종합하여 대표의 자격으로 국립국어원과 소통한다. 또한, 사업의 진행 방향과 일정에 차질이 없도록 사업 주최 기관인 국립국어원에 사업의 진행 척도를 주기적으로 보고하는 역할을 한다.

다음은 말뭉치구축팀이다. 데이터리의 오진영 이사가 팀장을 맡으며 구축 작업을 위해 단기적으로 고용한 20명의 작업자로 이루어져 있다. 말뭉치구축팀은 말뭉치 구축 작업을 위한 도구를 개발한다. 도구에는 오류 검증 및 개체명 부착을 위한 자동화 시스템과 작업자들이 개체명 부착 작업 시 사용하는 ‘Kane’이 포함된다. 또한, 작업자들을 교육하여 사업의 주된 작업인 개체명 부착 작업이 원활히 실행되도록 관리한다.

품질보증팀은 창원대학교 국어국문학과 정경재 교수를 팀장으로 주축을 이룬다. 품질보증팀에서는 개체명 부착 작업이 이루어진 말뭉치에 대해 최종 검증을 한다. 여기서 나온 오류들은 말뭉치구축팀으로 전달되어 수정작업이 이루어진다. 또한, 원문 자료나 원시 말뭉치 등을 검증하며 지침과 관련해 변경이 이루어져야 할 부분들을 작업한다.

마지막으로 기술보증팀에서는 개체명 구축과 검증에 이루어지는 도구를 개발하며, 작업 내내 사용될 웹의 환경을 구축한다. 또 외부의 접근 및 보안에 대한 관리를 맡아 사업이 철저한 보안 속에 이루어지도록 힘쓴다.

앞서 설명한 수행조직의 구조도와 사업 이행 작업자 명단은 다음과 같다.



<그림 8> 수행조직 구조도

□ 말뚝치 구축 작업 관리 방법

지침 교육 및 도구 사용 방법 안내를 위한 오리엔테이션을 진행하였다. 교육 후에는 배포 문서가 아닌 교육 문서를 먼저 배포하여 작업자들이 개체명 부착에 대한 이해가 이루어지도록 약 2주간의 적응 기간을 가진다. 충분한 지침 숙지와 어느 정도 작업이 적응되었다고 판단되었을 때 국립국어원에서 전달받은 원시 말뭉치 문서를 배포하여 본격적으로 작업을 진행한다.

작업 문서의 배포와 제출은 본 사업을 위해 개설한 비공개 웹을 통해서 진행한다. 관리자와 작업자의 가입만 허용한 비공개 형식의 웹을 통하여 관리자가 작업 문서를 배포하며 작업자는 완료한 문서제출을 하기도 하고 작업 시 발생하는 질문을 등록한다.

나, 너, 우리, 그 그리고, 자는, 당신, 초음향생, 골격생, 핵발, 입구, 이빨러야 복부 메르디나
#803 - opened 3 days ago by 조세훈 [#129](#)

대강 질문입니다
#802 - opened 2 days ago by 김성현 [#129](#)

목/박니돌목 두 관한 질문 있습니다.
#801 - opened 4 days ago by 진원호 [#129](#)

자유 사찰문자는
#800 - opened 4 days ago by 조세훈 [#129](#)

성형수술과 같이 수술을하는 대강대상에 포함되나요?
#797 - opened 4 days ago by 박상민 [#129](#)

황사/일타/이노출/가죽 등 관한 질문있습니다.
#800 - opened 4 days ago by 김성현 [#129](#)

대강 관한 질문입니다.
#795 - opened 4 days ago by 김해리 [#129](#)

이탈 레부타, 초음향생 핵
#804 - opened 4 days ago by 조세훈 [#129](#)

만용물과 APW 안가?
#802 - opened 4 days ago by 박상민 [#129](#)

스프즈 관한 행사가 아닌 대박망
#801 - opened 3 days ago by 김지희 [#129](#)

대강 관한 질문입니다
#801 - opened 3 days ago by 임아람 [#129](#)

[**실전 3차**] **E** 풀더 업로드
#205 - opened about 7 hours ago by 고은지 **익힘**

[**실전 3차**] **L** 풀더 업로드
#203 - opened about 10 hours ago by 양지혜 **익힘**

[**실전 3차**] **K** 풀더 업로드
#202 - opened about 10 hours ago by 어종의 **익힘**

[**실전3차**] **J** 풀더 업로드
#201 - opened about 11 hours ago by 천영우 **익힘**

[**실전3차**] **M** 풀더 업로드
#200 - opened about 11 hours ago by 제재영 **익힘**

[**실전3차**] **O** 풀더 업로드
#199 - opened about 12 hours ago by 연니영 **익힘**

[**실전 3차**] **F** 풀더 업로드
#198 - opened about 12 hours ago by 임아람 **익힘**

[**실전 3차**] **N** 풀더 업로드
#197 - opened about 13 hours ago by 안지섭 **익힘**

[**실전 3차**] **G** 풀더 업로드
#196 - opened about 14 hours ago by 김대형 **익힘**

〈그림 9〉 작업자들의 질문사항과 작업물 제출 페이지

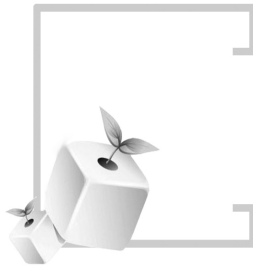
이에 대한 답변은 관리자와 품질 보증팀에서 작성한다. 등록된 질문 및 답변은 모든 작업자가 확인할 수 있도록 엑셀 파일로 정리된다. 또 기타 공지사항들을 실시간으로 공유하여 작업에 차질이 없도록 한다.

[illegible][illegible]

〈그림 10〉 작업자들 제출 현황 및 질문/답변 사항 엑셀 파일 정리 화면

□ 말뚱치 구축 수정·보완 작업 진행

말뚱치 구축팀에서 개체명 구축 작업이 완료되면 오류 검증을 위하여 품질 보증팀에 이를 전달한다. 품질 보증팀은 문서를 2차 검토하여 오류 사항을 수정한다. 오류 사항은 다시 말뚱치 구축팀에게 전달하여 작업자들에게 공지하고 이를 말뚱치구축팀에서 자동화 시스템에 적용하여 자동화 도구의 성능을 향상시킨다.



제 3 장

개체명 분석 말뭉치 구축 지침



제3장 개체명 분석 말뭉치 구축 지침

본 사업은 300만 어절의 개체명 말뭉치 구축과 함께 한국전자통신연구원(ETRI, 이하 ETRI)에서 만든 ‘세부분류 개체명 가이드라인 2018’을 정비하는 것을 목적으로 진행되었다. ETRI의 개체명 태그 세트는 15개의 대분류와 140여 개의 세분류로 구성되어 있다. 이의 일부를 제시하면 아래와 같다.

대분류	세분류	정의
PERSON(PS)	PS_NAME	사람 이름
LOCATION(LC)	LC_OTHERS	기타 장소
	LCP_COUNTRY	국가명
	LCP_PROVINCE	도, 주 지역명
	LCP_COUNTY	세부 행정구역명
	LCP_CITY	도시명
	LCP_CAPITALCITY	수도명
	LCG_RIVER	강, 호수, 연못
	LCG_OCEAN	해양, 바다
	LCG_BAY	반도, 만
	LCG_MOUNTAIN	산, 산맥, 능선, 고개
	LCG_ISLAND	섬, 제도
	LCG_CONTINENT	대륙
	LC_TOUR	관광명소
	LC_SPACE	천체 명칭
ORGANIZATION(OG)	OG_OTHERS	기타 기관/단체
	OGG_ECONOMY	경제 관련 기관/단체
	OGG_EDUCATION	교육 관련 기관/단체
	OGG_MILITARY	군사 기관/단체, 국방 기관
	OGG_MEDIA	미디어 기관/단체, 방송 관련 기관/기업
	OGG_SPORTS	스포츠 기관/단체
	OGG_ART	예술 기관/단체
	OGG_MEDICINE	의학/의료 기관/단체

	OGG_RELIGION	종교 기관/단체, 종파
	OGG_SCIENCE	과학 기관/단체
	OGG_LIBRARY	도서관 및 도서관 관련 기관/단체
	OGG_LAW	법률 기관/단체
	OGG_POLITICCS	정부/행정 기관, 공공 기관
	OGG_FOOD	음식 관련 업체/회사
	OGG_HOTEL	숙박 관련 업체
ARTIFACTS(AF)	AF_CULTURAL_ASSET	문화재(건물, 서적, 작품 등)
	AF_BUILDING	건축물/토목 건설물, 운동장 이름, 아파트, 다리, 등대, 분수
	AF_MUSICAL_INSTRUMENT	악기 명칭
	AF_ROAD	도로/철로 명칭
	AF_WEAPON	무기 명칭
	AF_TRANSPORT	교통수단, 운송수단, 놀이기구
	AF_WORKS	작품 세부 유형에 해당하지 않는 기타 작품명
	AF_WARES	상품/제품 이름

<표 1> ETRI의 개체명 분류 및 태그 세트 일부

본 사업에서는 ETRI에서 정리한 개체명 대분류와 그 태그를 따라 태깅하는 것으로 하였으며, 세분류는 구분하여 태깅하지 않았다. 본 사업의 개체명 분류와 태그 세트를 제시하면 아래의 표와 같다.

대분류	세분류	태그
인물	인명	PS
지역	지명, 국가, 도시, 수도, 바다 등	LC
기관	경제 기관, 교육 기관, 군사 기관 등	OG
인공물	문화재, 건축물, 도로, 작품명 등	AF
날짜	날짜 기간, 절기, 달(월), 계절 등	DT
시간	시간, 기간, 시각, 분, 초 등	TI
문명	민족, 종족, 제도, 언어, 직업 등	CV

동물	포유류, 조류, 파충류, 양서류 등	AM
식물	과일, 꽃, 나무, 풀 등	PT
수량	무게, 길이, 넓이, 개수, 온도 등	QT
학문 분야	철학, 의학, 예술, 사회과학 학과 등	FD
이론	철학 이론/사상, 예술 이론/양식 등	TR
사건	전쟁, 혁명, 스포츠 행사, 축제 등	EV
물질	금속, 암석, 화합물, 원소	MT
용어	색, 방향, 질병, 이메일 주소 등	TM

<표 2> 본 사업의 개체명 분류 및 태그

본 사업은 ETRI에서 만든 개체명 말뭉치 지침을 기본으로 하되, 이 지침을 수정·보완하는 것을 목표로 하였다. 이를 위해 본 사업팀에서는 다음과 같이 12차에 걸쳐 국어원과 지침에 대한 의견을 주고받았으며, 협의된 사항을 지침에 반영하는 작업을 진행하였다.

위와 같은 논의의 결과 수정·보완된 지침의 대략은 다음과 같다. 이 장에서는 중요한 사항만을 요약하여 언급하도록 한다.

1 대원칙 보완

개체명 말뭉치를 구축할 때에는 세 가지 문제에 대한 판단이 이루어져야 한다. 우선 텍스트에서 개체명의 대상을 파악하고, 이를 단독으로 태깅할 것인지 주변 단어와 함께 태깅할 것인지 판단하며, 어떤 태그를 부착할 것인지 결정해야 한다. 이와 같은 ‘태깅 대상 파악’, ‘태깅 단위 판단’, ‘태깅 범주 결정’을 개체명 태깅의 세 가지 문제라 할 수 있다. 기존의 지침은 그중 세 번째 문제, 즉 어떤 태그를 부착할 것인지에만 집중되어 있어 나머지 두 문제에 대한 지침이 불명확하고 소략하였다. 따라서 이번 지침 수정 작업에서는 우선 이를 보완하고자 하였다.

첫째, ‘개체명’이 무엇인지 재정의하여 태깅 대상의 혼란을 줄이고자 하였다. 기존의 ETRI 지침을 포함하여 개체명에 대한 여러 연구들에서 ‘개체명’은 ‘고유명사’를 대상으로 한다고 되어 있으나 실제 개체명으로 태깅되는 대상은 고유명사에 한정되지 않으며 많은 일반 명사를 포함한다. 개체명을 고유명사로 한정하면 인명, 지명, 작품명, 단체명 등 특정 범주에 대한 것으로 대상이 제한

되어 버려 말뭉치를 구축하는 본래의 의도와 결과물이 달라지는 문제가 있기 때문이다. 이에 따라 고유명사뿐 아니라 특정 범주의 세부 분류명도 개체명에 포함됨을 지침에 명시하였다. 예를 들어 AM과 PT에는 동물과 식물의 다양한 종이 개체명 대상이 되며, MT에는 원소나 금속, 암석의 구체적인 유형이 개체명의 대상이 되는 등이다. 그러나 4장의 ‘남은 문제’에서 언급될 바와 같이 이 문제는 15가지 개체명 범주마다 서로 다른 기준을 가지고 있어 아직 보완할 부분이 많이 남아 있다.

둘째, 태깅 단위에 대해서는 ‘다. 태깅 단위와 태깅 범주 문제’에서 상세화하였다. 기존 지침에서는 단어 단위로 작게 태깅해야 하는 유형과 구 단위로 통합하여 태깅해야 하는 유형을 몇몇 사례만으로 보이고 있어 실제 말뭉치에서 확인되는 다양한 예에 적용하는 데 한계가 있었다. 이를 일반화된 기술로 수정하고 ‘LC + LC’, ‘LC + AF’, ‘LC + OG’, ‘OG + AF’, ‘DT, QT + CV, EV’ 등과 같이 자주 등장하는 사례를 태그 조합별로 유형화하여 태깅 단위를 설명하였다.

뿐만 아니라 태깅 단위와 관련해서는 복합어의 내부 요소를 태깅할 것인지에 대한 문제도 정리될 필요가 있어 이에 대한 지침을 추가하였다. 복합어는 구성 요소와는 별개의 개체를 지시하게 되므로 복합어가 개체명인 경우 복합어 전체를 개체명으로 태깅하며, 복합어가 개체명이 아닌 경우는 내부 요소가 개체명이더라도 별도로 태깅하지 않는 것을 대원칙으로 삼았다. 그러나 선행 어기의 의미에 영향을 주지 않는 접사 ‘-들’ 결합형 등 복합어 내부 요소를 태깅하는 일부 예외가 존재한다. 이에 대해서는 ‘다. 태깅 단위와 태깅 범주 문제’에서 상론하였다.

셋째, 태깅 범주에 대해서는 기존 지침에서도 비교적 상세하게 정리되어 있었다. 그러나 하나의 단어 혹은 구가 둘 이상의 의미 영역에 동시에 속하여서 중복 태깅이 가능한 경우, 어떤 태그를 부착할 것인지에 대해서는 정리가 미흡하였다. 따라서 이에 대해 가능한 유형을 모두 정리하여 두 가지로 분류하여 기준을 세웠다. 우선 개체명이 문맥과 관계없이 ‘본유적’으로 둘 이상의 범주에 속하는 경우는 지침에서 우선순위를 명시하였다. 예를 들어 ‘경포대’는 관광지로서 LC_TOUR에 속하며, 건축물로서 AF_BUILDING에 속하고 문화재로서 AF_CULTURAL_ASSET에 속하기도 한다. 따라서 이러한 경우 ‘AF_BUILDING <

AF_CULTURAL_ASSET < LC_TOUR' 라는 우선순위에 따라 LC_TOUR로 태깅한다.

반면 개체명이 문맥에 따라 하나의 범주에 속하는 것으로 이해되는 경우에는 문맥에 따라 태깅하는 것을 우선 원칙으로 삼고 문맥이 중의적인 경우에만 우선순위를 적용하는 것으로 구분하였다. 예를 들어 ‘미국, 일본’ 과 같은 국가명의 경우, 장소로서 LC이면서 국가로서 하나의 단체인 OG에 속하는데, 이는 문맥에 따라 구분이 가능한 경우가 많으므로 그에 따라 LC 혹은 OG로 태깅되 중의적인 경우에만 우선순위에 따라 LC로 태깅하도록 하였다.

2 전체 부류에 적용되는 일부 지침 추가

개체명 말뭉치 구축의 대원칙이라고 할 수는 없으나 태깅 시 중요하게 고려해야 할 사항으로서, 15가지 개체명 부류에 두루 적용되는 몇 가지 지침을 신설하였다.

첫째, 구어 말뭉치 태깅 원칙을 ‘라. 기타 / 3. 구어 개체명 태깅’ 부분에 추가하였다. 기존의 ETRI 지침은 문어 말뭉치를 대상으로 한 것이어서 구어 말뭉치에서만 나타나는 특수한 형태를 태깅하는 방안을 포함하고 있지 않았다. 구어 말뭉치에서는 발음 오류로 인해 개체명의 어형이 잘못된 형태로 나타나기도 하며, 체언과 조사 혹은 (분절음으로 실현되지 않은 서술격조사에 후행한) 어미가 결합하여 하나의 음절로 실현되기도 한다. 이들은 모두 개체명으로 태깅하되 후자의 경우 후처리로 체언과 문법 형태소를 구분하는 방식을 취하기로 하였다.

둘째, 개체명 뒤에 괄호와 함께 원어 정보가 제시된 경우, 원어 정보를 태깅하는 방식에 대한 지침을 새로 추가하였다. 이는 ‘라. 기타 / 2. 기타 논의 사항’에 정리되어 있다. 원어 정보는 선행하는 개체명과 동일한 정보를 나타낼 때에는 태깅하지 않으나, 준말 등으로 나올 경우 괄호 밖의 개체명과 다른 형태로서 하나의 개체명이 되는 것이므로 태깅하는 것을 큰 원칙으로 삼았다.

셋째, 개체명을 태깅할 때 문맥을 고려해야 하는지, 해당 표현이 지닌 본래적 의미만을 고려하여 태깅해야 하는지에 대한 고민의 일부를 지침으로 마련하였다. 우선 하나의 텍스트에서 앞에 나온 단어 혹은 구가 텍스트의 뒷부분에서는 어형의 일부가 절단된 상태로 쓰인 경우(즉 축소 지칭된 경우), 텍스트 내 다른

문장이나 문맥을 고려해야만 어형이 복원된다면 홀로는 그 의미를 지니지 못한다고 보고 절단된 단어 자체의 의미 범주에 준하여 태깅하도록 하였다. 반대로 텍스트 내 다른 문장이나 문맥을 고려하지 않고도 축소 지칭된 어형이 본래 어형의 의미를 지닐 수 있는 경우는 복원된 형태에 준하여 태깅하도록 하였다.

예를 들어 ‘과천시립미술관’이 텍스트의 뒷부분에서 ‘미술관’만으로 쓰인 경우, ‘미술관’은 문맥 없이 ‘과천시립미술관’의 의미를 지니지 못하며 홀로는 구체적인 지시물을 가리키지 못하므로 태깅하지 않는다. 반면 ‘FC 서울’이 텍스트의 뒷부분에서 ‘서울’로만 쓰인 경우 “어제 서울이 수원을 이겼다”와 같이 ‘서울’만으로 ‘FC 서울’을 의미할 수 있으므로 ‘FC 서울’에 준하여 태깅한다.

뿐만 아니라 어떠한 단어가 ‘비유적’인 의미로 쓰인 경우, 비유인지 아닌지를 고려하지 않고 본래의 개체명 부류대로 태깅하기로 하였다. 예를 들어 “앓던 이가 빠졌다”라는 표현에서 ‘앓던 이’는 “걱정거리”를 의미한다. 따라서 문맥 의미를 따르면 이는 개체명이 아니다. 그러나 문맥 의미를 고려하지 않고 ‘이’가 지닌 신체 부위로서의 본래 의미에 따라 AM_PART로 태깅한다. 이러한 문제에 대해서는 ‘라. 기타 / 2. 기타 논의 사항’에 정리되어 있다.

즉 텍스트의 뒷부분에서 축소 지칭된 경우와 비유적으로 사용된 경우는 문맥 의미를 고려하지 않고 해당 단어나 구가 갖는 본래의 의미에 준하여 태깅하도록 하였다. 그러나 그 외의 경우에 문맥에 따른 의미 변화를 얼마나 인정하여 태깅해야 할 것인지 하는 문제는 앞으로의 지침에 더 보완되어야 한다. 이에 대해서는 4장의 남은 문제에서 부연할 것이다.

3 부류별 지침 수정

그 외에 각 부류별로 세부 지침이 수정되었다. 여기에서 모두 나열할 수는 없으므로 대표적인 사례 몇 가지만 정리하도록 한다.

① STUDY FIELD(학문 분야 및 학과)

학문 분야의 경우 FD_ART는 예술 관련 학문 분야 외에도 예술의 분류, 장르가 태깅 대상에 포함이 된다는 점에서 다른 분야와는 차이가 있다. 이러한 기준

지침을 수용하되, 예술의 ‘장르’가 아닌 예술 활동의 ‘결과물(작품)’은 FD_ART로 태깅하지 않는다는 지침을 추가하였다. 예술 작품은 구체적인 작품명으로 나올 경우 학문 분야(FD)가 아닌 인공물(AF)이기 때문이다.

② ARTIFACTS(인공물)

구체적인 상품명을 대상으로 하는 AF_WARES는 다른 개체명 부류와 중복되는 경우 다른 개체명 부류에 우선순위를 부여하는 지침이 있었다. 그러나 ‘아이폰’이나 ‘17차’와 같은 구체적인 상품명은 AF_WARES로 태깅하며 ‘휴대폰’이나 ‘녹차’와 같은 분류 명칭은 TRM_HARDWARE나 CV_DRINK 등 여타의 부류로 태깅하는 것으로 차별화하였다.

③ ORGANIZATION(기관 및 조직)과 LOCATION(지명)

국가명이나 행정구역명은 맥락에 따라 OG가 될 수도 있고 LC가 될 수도 있다. 이 경우 문장에서 어떠한 의미로 사용되었는지에 따라 태깅을 달리한다는 지침을 새로 추가하였다. 이때 판단 기준은 서술어를 중심으로 한 문장 의미에 따라 결정한다. 또한 ‘명사(국가/도시/행정구역명) + 명사’ 구성에서(예: 서울 시장) 선행하는 명사는 LC로도 OG로도 볼 수 있으므로 일괄 LC로 태깅한다.

④ CIVILIZATION(문명)

CV 태깅 대상에는 의, 식, 주 등 다양한 요소가 포함되어 있다. 그중 CV_FOOD는 AM 또는 PT에 속할 경우 AM, PT에 우선순위를 부여하는 지침이 있었다. 그러나 문맥상 음식 재료명으로 쓰였다고 판단되는 경우는 CV로 태깅하는 것으로 지침을 수정하였다. CV_CLOTHING의 경우 모자 종류는 태깅하지 않는다는 지침이 있었으나, 모자 역시 의류의 일종으로 볼 수 있으므로 태깅 대상에 포함하는 것으로 지침을 수정하였다.

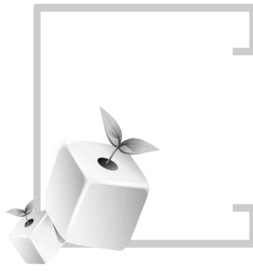
⑤ DATE(날짜), TIME(시간), QUANTITY(수량)

날짜, 시간, 수량의 경우 기존 지침에 제시된 사례를 살펴보면, 해당 수량 표현을 구체화시키는 ‘보조사, 접사, 단위성 의존명사’ 등은 함께 태깅되는 경향을 볼 수 있었으나 지침에 조항으로서 명시되어 있지 않았다. 따라서 이들을 하나

의 덩어리로 통합하여 태깅한다는 지침을 추가하였다.

또한 기존 지침에서는 ‘첫’을 QT로 태깅하도록 되어 있었으나 개체명은 체언(구)만을 대상으로 한다는 큰 지침에 근거하여 ‘첫’은 홀로 태깅하지 않도록 하였으며, ‘첫 + 단위명사’와 ‘첫 + 일반명사’에서 일반명사가 낱자 관련 명사일 때에는 함께 덩어리로 태깅하는 것으로 지침을 수정하였다.

그 외에 여기에서 모두 상론할 수는 없으나, 구체적인 사례로 나열된 지침 중 일반화된 기술로 변경이 가능한 것은 모두 수정하였으며, 기타 불명확하게 기술된 지침은 명확하게 다듬었다. 이를 통해 이후 이 지침을 이용하여 개체명 말뭉치를 구축하려는 연구자가 오해 없이 지침을 사용할 수 있도록 도왔다. 그럼에도 이번에 수정된 지침은 아직 완벽하지 않다. 이는 사업 주최 기관인 국립국어원에서 기존 ETRI 지침의 큰 틀을 유지하고자 하는 방침을 유지하였으며, 또한 개체명의 하위 부류가 본유적으로 불명확한 경계를 지니고 있는 데에서 말미암은 것이다. 남은 문제에 대해서는 4장에서 다시 언급하도록 한다.



제 4 장

결론



제4장 결론

1 사업 요약

국립국어원이 주관하고 기업 데이터리(Data.ly)와 공동참여기관 창원대학교가 함께 참여하여 진행한 ‘개체명 분석 말뭉치 구축’ 사업은 2019년 7월 22일부터 시작하여 2019년 11월 22일까지 약 4개월간 진행되었다. 사업의 주 내용은 원시 말뭉치 300만 어절에 대한 개체명 분석과 말뭉치 부착 작업을 오류율 5% 이내로 완료하는 것과 작업에 쓰인 ETRI의 ‘세부분류 개체명 가이드라인 2018’의 내용을 수정 및 보완하는 것이었으며, 완료된 작업물은 민간에 배포되어 관련 연구 및 사업에 이바지하는 것을 목표로 삼았다.

사업 수행 기관인 데이터리와 창원대학교는 각자 지닌 전문적인 지식에 따라 담당 역할을 나누었고 크게 ‘말뭉치 구축팀’과 ‘품질 보증팀’으로 분류하였다. 말뭉치 구축팀에서는 구축 프로그램 ‘kane’을 이용한 개체명 부착 작업과 자동화 시스템을 이용한 오류검출 및 수정작업과 인력 관리를 맡았으며 품질 보증팀에서는 국어학적 지식을 지닌 전문가가 지침에 대해 검토하고 문제점에 대해 보완하는 작업과 완료된 작업물들에 대해 다시 한번 확인하는 작업을 거쳐 오류 사항들을 잡아냈다.

작업 진행 과정에서는 말뭉치 구축팀에서 자체 개발한 개체명 구축 프로그램 ‘kane’과 자동화 시스템을 이용한 말뭉치 정책 적합 여부 작업을 통해 인력과 컴퓨터 기술을 융합하여 작업을 진행함으로써 빠르고 정확하게 작업을 진행할 수 있었다.

지침 수정 및 보완의 과정에서는 ‘개체명’의 정의를 명확히 하고 태깅의 단위와 범위에 대하여 상세화하는 작업을 진행하였다. 그 외에도 15가지 개체명 부류에 적용되는 몇 가지 지침을 신설하거나 지침의 세부적인 내용을 수정함으로써 지침의 내용이 더 완전해지도록 보완하였다.

본 사업을 진행하며 참여 기관인 데이터리와 창원대학교는 주관기관인 국립국어원에 개체명 분석과 말뭉치 구축 작업을 마친 원시 말뭉치 300만 어절과 수정 및 보완이 이루어진 지침, 또 사업의 진행내용을 간략히 기술한 최종 보고서를 납품하며 본 사업을 마무리한다.

국립국어원이 주관한 ‘개체명 분석 말뭉치 구축’ 사업이 성공적으로 마무리되었다. 그러나 본 사업에서 개체명 분석 말뭉치 구축에 관한 모든 문제를 해결한 것은 아니며, 여전히 아쉬운 점은 남아 있다. 추후 진행될 관련 사업에 도움이 되고자 남은 문제에 대해 정리하고자 한다.

개체명 분석 말뭉치를 구축할 때 부딪치는 가장 어려운 점은 ‘개체명’이 무엇이며, 개체명 말뭉치 구축이 무엇을 목적으로 하는지, 정의와 지향점이 아직 선명하지 않다는 점이다. 앞서도 언급했듯이 기존의 ETRI 지침이나 개체명에 관한 기존의 연구들에서 개체명은 ‘고유명사’를 대상으로 한다고 되어 있다. 그러나 실제 태깅 대상은 많은 일반명사를 포함하는데, 일반명사를 어디까지 태깅해야 하는지 그 경계가 분명치 않다. 개체명 분석 말뭉치의 학술적·산업적 용도를 고려하여 ‘개체명’ 개념에 대한 논의가 학계에서 더 활발히 이루어질 필요가 있다.

원시 말뭉치에 비해 주석(태깅) 말뭉치를 구축하는 사업은 상대적으로 일관성 있고 명확한 지침, 고성능·고효율 프로그램, 국어학 지식을 지니고 있으며 지침을 숙지한 작업자가 갖추어져야 성공적으로 완수될 수 있다. 이 중 가장 우선적으로 마련되어야 하는 것이 ‘말뭉치 구축 지침’이다. 본 사업은 말뭉치 구축과 지침 정비를 동시에 수행하도록 계획되어, 300만 어절의 개체명 말뭉치를 구축하면서 지침을 수정하는 작업이 병행되었다. 그로 인해 사업의 종료 시점까지 지침이 변경되어 태깅이 끝난 전체 데이터를 다시 수정해야 하는 작업이 수차례 반복되었다. 이는 작업자의 혼란뿐 아니라 기계 학습에도 오류를 야기했다. 향후의 개체명 말뭉치 구축 사업에서는 보다 정비된, 완성도가 높은 지침이 배포된 상태로 말뭉치 구축이 시작되어야 할 것이다. 이를 위해 이번에 수정·보완된 개체명 말뭉치 구축 지침에 남아 있는 문제를 정리하고자 한다.

첫째는 개체명 말뭉치 태깅 대상의 특정성 경계 문제이다. 개체명은 모든 고유명사와 지침에서 규정하고 있는 일부 일반명사를 포함한다. 일반명사의 경우, 대체로 추상적이고 포괄적인 개념의 상의어는 제하고 있는데 그 경계가 명확하지 않으며, 부류마다, 부류 안에서도 사례마다 차이가 있다. 예를 들어 EVENT의 경우 EV_ACTIVITY와 EV_WAR_REVOLUTION은 구체적인 사건만을 대상으로 하나

EV_SPORTS와 EV_FESTIVAL은 ‘마라톤 대회’나 ‘연등축제’ 같은 일반 유형도 태깅 대상으로 삼는다. 뿐만 아니라 EV_OTHERS는 ‘정상회담, 투표, 선거, 시험·고사’의 경우만 일반 유형을 태깅하며 그 외의 유형은 구체적인 사건만을 태깅하도록 하고 있다. 또 다른 사례로, OGG_EDUCATION의 경우 ‘훈련소’만 일반 유형을 태깅 대상으로 삼고 있는데, 이유를 알 수 없다. OGG_POSITICS에서 ‘시, 군, 구’가 단독으로 나오면 특정한 기관을 의미하지 못하는데도 태깅하도록 한 것도 일관성이 없는 사례이다. 이처럼 명확한 이유 없이 태깅 대상의 특정성-추상성 정도에 차이가 있는 사례가 수정 지침에 여전히 남아 있는데, 이는 개체명의 정의와 개체명 말뭉치 구축의 지향점에 따라 일관성 있게 다듬어져야 할 것이다.

덧붙여 일반명사를 어디까지 태깅할 것이냐 하는 이해를 돕기 위해서는 지침에 제시된 예들이 다채로워져야 한다. 현재는 가장 전형적인 예들만을 제시하고 있는데, 추상성·특정성의 정도가 서로 다른 사례를 제시하여 어디까지 태깅 대상으로 삼고 있는지를 보여 주어야 한다.

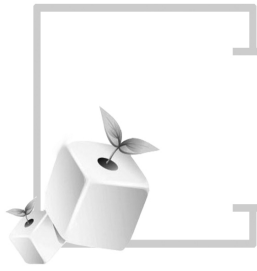
둘째는 개체명에 태깅 범주를 부여할 때 문맥을 얼마나 고려할 것인가 하는 문제이다. 현 지침에서는 하나의 단어나 구가 ‘본유적으로’ 두 가지 범주에 속할 때(예: ‘시금치’는 CV_FOOD이면서 PT_GRASS임) 문맥을 고려하여 태깅하도록 하였으나, 만약 어떤 단어나 구가 문맥에 따라 ‘일시적으로’ 다른 범주에 속하는 것으로 쓰였을 때 어떤 범주로 태깅해야 하는지는 명시되어 있지 않다. 예를 들어, ‘기타’는 악기이므로 AF_MUSICAL_INSTRUMENT에 속하는데 문맥에 따라 밴드의 포지션을 의미할 수 있다. 이 경우 CV_POSITION으로 태깅해야 할 것인가 하는 문제가 지침에 기술되어야 한다. 본 사업팀에서는 사업 주최 기관인 국립국어원의 의견에 따라 문맥을 고려하여 태깅하였으나, 이것이 개체명 말뭉치의 본래 지향점에 부합하는지도 재고가 필요하다.

세 번째는 복합어 내부 요소의 태깅 여부 문제이다. 앞서도 언급하였듯이 이번 사업의 지침 보완 작업 과정에서 본래 지침에 명시되어 있지 않았던, 복합어 내부 요소를 태깅하지 않는다는 지침이 추가되었다. 복합어는 전체가 별개의 개체를 지시하므로 복합어 전체가 개체명일 경우에만 태깅하고 복합어 전체가 개체명이 아닌 경우에는 내부에 개체명이 있더라도 태깅하지 않기로 한 것이다. 그러나 생산적인 접사가 결합한 파생어(예: 한국적(韓國的), 16강행, 반아베파)와 생

산적으로 합성어를 형성하는 후행 요소가 결합하여 의미가 투명한 합성어(예: 삼겹살집, 고구마밭)도 태깅하지 않는다면 ‘한국, 16강, 아베, 삼겹살, 고구마’와 같은 많은 개체명을 태깅하지 못하는 결과를 낳는다. 복합어 내부 요소의 태깅에 대해서는 현 지침에 명확하게 기술이 되어 있으나 향후 재고를 요한다.

네 번째, 개별 사례가 나열될 뿐 일반화된 진술로 기술할 수 없는 지침이 여전히 남아 있다. 직위·직책 명칭 등을 대상으로 하는 CV_POSITION이 대표적인 사례이다. 현 지침에서는 직위나 직책뿐 아니라, “사람들이 가질 수 있는 특정한 역할”도 태깅 대상이 된다고 보고 “일반 사람이 텔레비전 프로그램을 보는 역할을 수행함으로 인해서” ‘시청자’가 개체명이 될 수 있다고 보고 있다. 이에 따라 직위·직책이 아닌 ‘피서객, 포로, 유부녀, 미혼모, 재일동포, 대리인’ 등과 지침에는 명시되지 않았으나 ‘사상자, 피해자’ 등도 모두 태깅 대상에 포함되었다. 반면 ‘처녀, 아낙네, 멘토’ 등은 태깅 대상이 아닌 것으로 명시하고 있다. ‘유부녀, 미혼모’가 어떠한 ‘역할’을 수행하는지 알기 어려울 뿐 아니라 ‘처녀, 아낙네’와의 차이도 불명확하다. 이 지침이 대규모 말뭉치를 구축하는 데 이용되기 위해서는 개별 사례의 나열이 아닌 일반화된 진술로 기술되어 있어야 예측하지 못했던 용례에 대한 처리도 가능할 것이다. 한편으로 사례의 나열로밖에 기술될 수 없는 지침은 문제가 있는 것이므로 과감한 수정이 필요하다.

마지막은 각 범주 안의 세부 사례에 대한 것이다. STUDY_FIELD에서는 학문 분야의 분류 체계에 ‘인문 과학’이 누락되어 있으며 ‘철학’만이 독자적인 분류로 구분되어 있다. 그 결과 ‘언어학’은 FD_SOCIAL_SCIENCE에 포함되며 ‘국어국문학’은 FD_OTHERS에 포함되어 있다. ARTIFACTS에 ‘무형문화재’ 종류가 포함되는 것도 문제이다. AF_CULTURAL_ASSET에 유형문화재가 포함되므로 그와의 유사성으로 무형문화재도 이에 포함되었으나 엄밀한 의미에서 무형문화재는 ‘인공물’이 아니다. ‘조약, 협정, 협약’을 OGG로 태깅하도록 한 것도 분류 체계상 적절하지 않은 것으로 보인다. ‘회의’가 EV에 속하는 것과 대비가 되는 부분이다. 이러한 개체명 분류 체계는 장기적인 관점에서 수정 여부를 검토할 필요가 있다.



부록

개체명 분석 말뭉치 구축 지침 2019



국립국어원

이 지침은 ETRI 지침을 기반으로 하여 재가공된 것입니다.

무단 복제, 수정 및 배포를 금합니다

목 차

가. 개요	9
나. 세부분류 개체명 정의	19
1. PERSON	19
1.1 PS_NAME	19
2. STUDY_FIELD	21
2.1 FD_SCIENCE	22
2.2 FD_SOCIAL_SCIENCE	23
2.3 FD_MEDICINE	24
2.4 FD_ART	25
2.5 FD_PHILOSOPHY	26
2.6 FD_OTHERS	26
3. THEORY	27
3.1 TR_SCIENCE	27
3.2 TR_SOCIAL_SCIENCE	29
3.3 TR_ART	30
3.4 TR_PHILOSOPHY	30
3.5 TR_MEDICINE	31
3.6 TR_OTHERS	32
4. ARTIFACTS	33
4.1 AF_CULTURAL_ASSET	33
4.2 AF_BUILDING	34
4.3 AF_MUSICAL_INSTRUMENT	35
4.4 AF_ROAD	35
4.5 AF_WEAPON	36
4.6 AF_TRANSPORT	36
4.7 AF_WORKS	37
4.8 AF_WARES	44
5. ORGANIZATION	45
5.1 OGG_ECONOMY	45
5.2 OGG_EDUCATION	46
5.3 OGG_MILITARY	48
5.4 OGG_MEDIA	48
5.5 OGG_SPORTS	49

5.6 OGG_ART	49
5.7 OGG_MEDICINE.....	50
5.8 OGG_RELIGION	50
5.9 OGG_SCIENCE	51
5.10 OGG_LIBRARY.....	52
5.11 OGG_LAW.....	52
5.12 OGG_POLITICS.....	53
5.13 OGG_FOOD	55
5.14 OGG_HOTEL.....	55
5.15 OG_OTHERS.....	56
6. LOCATION	57
6.1 LCP_COUNTRY.....	57
6.2 LCP_PROVINCE	58
6.3 LCP_COUNTY	58
6.4 LCP_CITY	59
6.5 LCP_CAPITALCITY.....	59
6.6 LCG_RIVER	60
6.7 LCG_OCEAN	60
6.8 LCG_BAY	61
6.9 LCG_MOUNTAIN.....	61
6.10 LCG_ISLAND.....	62
6.11 LCG_CONTINENT.....	62
6.12 LC_TOUR.....	63
6.13 LC_SPACE	63
6.14 LC_OTHERS	64
7. CIVILIZATION	66
7.1 CV_NAME	66
7.2 CV_TRIBE	66
7.3 CV_SPORTS.....	67
7.4 CV_SPORTS_INST.....	68
7.5 CV_POLICY	68
7.6 CV_TAX.....	69
7.7 CV_FUNDS	69
7.8 CV_LANGUAGE	70
7.9 CV_BUILDING_TYPE	70
7.10 CV_FOOD.....	71
7.11 CV_DRINK.....	71
7.12 CV_CLOTHING	72
7.13 CV_POSITION	72
7.14 CV_RELATION	74

7.15 CV_OCCUPATION	74
7.16 CV_CURRENCY.....	75
7.17 CV_PRIZE	75
7.18 CV_LAW	76
7.19 CV_FOOD_STYLE.....	78
8. DATE.....	79
8.1 DT_DURATION	79
8.2 DT_DAY	79
8.3 DT_MONTH.....	80
8.4 DT_YEAR.....	80
8.5 DT_SEASON.....	80
8.6 DT_GEOAGE.	81
8.7 DT_DYNASTY	81
8.8 DT_OTHERS	82
9. TIME.....	83
9.1 TI_DURATION.....	83
9.2 TI_HOUR.....	83
9.3 TI_MINUTE.....	84
9.4 TI_SECOND	84
9.5 TI_OTHERS.....	84
10. QUANTITY	85
10.1 QT_AGE	85
10.2 QT_SIZE	85
10.3 QT_LENGTH.....	85
10.4 QT_COUNT.....	86
10.5 QT_MAN_COUNT	86
10.6 QT_WEIGHT	87
10.7 QT_PERCENTAGE	87
10.8 QT_SPEED.....	87
10.9 QT_TEMPERATURE	88
10.10 QT_VOLUME.....	88
10.11 QT_ORDER.....	88
10.12 QT_PRICE.....	89
10.13 QT_PHONE.....	89
10.14 QT_SPORTS	89
10.15 QT_CHANNEL	90
10.16 QT_ALBUM.....	90
10.17 QT_ZIPCODE	91
10.18 QT_OTHERS	91

11. EVENT	92
11.1 EV_ACTIVITY	92
11.2 EV_WAR_REVOLUTION.....	92
11.3 EV_SPORTS	93
11.4 EV_FESTIVAL.....	94
11.5 EV_OTHERS	95
12. ANIMAL.....	96
12.1 AM_INSECT	96
12.2 AM_BIRD	96
12.3 AM_FISH.....	97
12.4 AM_MAMMALIA.....	97
12.5 AM_AMPHIBIA.....	97
12.6 AM_REPTILIA.....	98
12.7 AM_TYPE	98
12.8 AM_PART.....	99
12.9 AM_OTHERS.....	99
13.PLANT	100
13.1 PT_FRUIT.....	100
13.2 PT_FLOWER.....	100
13.3 PT_TREE.....	101
13.4 PT_GRASS.....	101
13.5 PT_TYPE	102
13.6 PT_PART.....	102
13.7 PT_OTHERS.....	102
14. MATERIAL.....	103
14.1 MT_ELEMENT.....	103
14.2 MT_METAL.....	103
14.3 MT_ROCK.....	104
14.4 MT_CHEMICAL.....	104
15.TERM	107
15.1 TM_COLOR.....	107
15.2 TM_DIRECTION.....	107
15.3 TM_CLIMATE.....	107
15.4 TM_SHAPE	108
15.5 TM_CELL_TISSUE	108
15.6 TMM_DISEASE.....	109
15.7 TMM_DRUG	109
15.8 TMI_HW	110

15.9 TMI_SW.....111

15.10 TMI_SITE.112

15.11 TMI_EMAIL.....113

15.12 TMI_MODEL.....113

15.13 TMI_SERVICE.....113

15.14 TMI_PROJECT.....114

15.15 TMIG_GENRE.....114

15.16 TM_SPORTS.....114

다. 태깅 단위와 태깅 범주 문제.....	116
1. 태깅 단위.....	116
1.1 구 단위의 태깅	116
1.2 복합어 내부 태깅	118
2. 유일 태깅 원칙.....	119
 라. 기타	121
1. 개체명 인식 성능 평가.....	121
2. 기타 논의 사항	122
3. 구어 개체명 태깅	124

가. 개요

본 문서는 개체명 분석 말뭉치 구축 지침으로, 개체명 각 부류의 정의와 대표 사례 및 태깅 정책, 태깅 시 고려해야 하는 사항, 예상되는 문제점 등에 대해 기술하고 있다.

개체명의 대분류는 15 가지로, PERSON(인명), STUDY_FIELD(학문분야 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. 15 가지의 대분류는 또다시 각각 복수의 하위 범주로 세분화되어 있는데, 이는 말뭉치의 구축 목적에 따라 15 가지 범주에 속하는 개체명의 유형과 범위를 이해하는 데 사용될 수도 있고, 하위 범주를 모두 구분한 태깅 말뭉치를 구축하는 데 이용될 수도 있다. 아래는 세부 분류 개체명의 목록이다.

대분류	세분류	정의
PERSON (PS)	PS_NAME	-사람이름
LOCATIO N (LC)	LC_OTHERS	- LC 계열의 세부유형이 아닌 기타 장소 - 예) 지하철/기차역, 공항, 버스터미널, IC, 시장, 발전소
	LCP_COUNTRY	- 국가명 - 예) 한국, 미국, 일본, 중국, 동독
	LCP_PROVINCE	-도,주지역명 - 예) 제주도, 경기도, 충청도, 강원도, 전라도, 경상도, 텍사스주
	LCP_COUNTY	- 군, 면, 읍, 리, 동 등과 같은 세부 행정구역명 - 예) 영월군, 도암면, 횡계리, 노암동, 상무우미아트빌 정보화마을
	LCP_CITY	- 도시명 - 예) 강릉, 부산, 태백, 뉴욕, 맨체스터, 덴버, 요코하마, 이스탄불
	LCP_CAPITALCITY	- 수도명 - 예) 서울, 동경, 북경, 베이징, 런던, 워싱턴
	LCG_RIVER	- 강, 호수, 연못 - 예) 섬진강, 충주호, 관곡지
	LCG_OCEAN	- 해양, 바다 - 예) 남극해, 북극해, 동해, 태평양, 산호해
	LCG_BAY	-반도,만 - 예) 도쿄만, 라몬만, 라오둥반도, 변산반도
	LCG_MOUNTAIN	- 산, 산맥, 능선, 고개/재, 봉 - 예) 라우산, 태백산맥, 차령산맥, 계룡산, 공룡능선, 박달

	LCG_ISLAND	-섬,제도 - 예) 나이팅게일섬, 로스제도, 백령도
	LCG_CONTINENT	- 대륙 - 예) 아시아, 아프리카, 아메리카, 오세아니아, 유라시아, 유럽
	LC_TOUR	- 관광명소 - 예) 온천, 해변, 관광지+해안, 식물원, 수목원, 정원, 공원
	LC_SPACE	-천체명칭 - 예) 항성, 행성, 위성, 유성, 별자리, 성운
ORGANIZATION (OG)	OG_OTHERS	- 기타 기관/단체 - 예) 타진요, 붉은악마
	OGG_ECONOMY	- 경제 관련 기관/단체
	OGG_EDUCATION	- 교육 기관/단체, 교육관련 기관 - 예) 학교, 학원, 어린이집, 훈련소, 학과
	OGG_MILITARY	- 군사 기관/단체 및 유형, 국방기관 - 예) 해군, 공군, 육군, 인민군, 해군기지, 국방품질관리소
	OGG_MEDIA	- 미디어 기관/단체, 방송관련 기관/기업 - 예) 방송국, 신문사
	OGG_SPORTS	- 스포츠 기관/단체 - 예) 국민체육진흥심의위원회, 국기원, 국군체육부대
	OGG_ART	- 예술 기관/단체 - 예) 공연장, 전시장, 박물관, 미술관, 극장, 영화관
	OGG_MEDICINE	- 의학/의료 기관/단체 - 예) 병원, 약국, 세계보건기구, 국립재활원, 경희의료원
	OGG_RELIGION	- 종교 기관/단체, 종파 포함 - 예) 증산도, 증산교, 기독교, 불교, 대한성공회, 한국기독교장로회
	OGG_SCIENCE	- 과학 기관/단체 - 예) 연구소/연구원, 국제전기통신연합, 국립국어연구원
	OGG_LIBRARY	- 도서관 및 도서관 관련 기관/단체 - 예) 국립중앙도서관, 국회도서관
	OGG_LAW	- 법률 기관/단체 - 예) 광주지방법원, 대한법무사협회, 저작권심의조정위원회
	OGG_POLITICS	- 정부/행정 기관, 공공기관, 정치기관 - 예) 정부/행정 기관: 미래부, 정보통신부, 우체국, 경찰서
	OGG_FOOD	- 음식 관련 업체/회사 - 예) 맥도날드, 김밥천국, 롯데리아, 맘스터치

	OGG_HOTEL	- 숙박관련업체 - 예) 호텔, 모텔, 민박, 펜션, 콘도, 리조트
ARTIFACT S (AF)	AF_CULTURAL_ASSET	- 건물/서적/작품 등 문화재 - 예) 경복궁, 서산마애삼존불상
	AF_BUILDING	- 건축물/토목건설물, 운동장이름, 아파트, 다리, 등대, 분수
	AF_MUSICAL_INSTRUMENT	- 악기 명칭 - 예) 기타, 캐스터네츠, 편경, 팽가리, 대금, 해금
	AF_ROAD	- 도로/철로 명칭 - 예) 세종로, 올림픽대로, 충북선, 경부선
	AF_WEAPON	- 무기명칭 - 예) 기관단총, 시한폭탄, 곡사포, 고정포, 탄도미사일, 도끼, 칼
	AF_TRANSPORT	- 교통수단/자동차/선박 모델 및 유형, 운송수단, 놀이기구 - 예) 갈라테아호, 경비정, 벤츠, 자기부상열차
	AF_WORKS	- 작품 세부유형에 해당하지 않는 기타 작품명 - 예) 뽀빠이 영철영어, 중종실록, 고려사
	AF_WARES	- 상품/제품 이름 - 예) 화장품, 운동화 이름, 여행상품
DATE (DT)	DT_OTHERS	- 세분류 외의 날짜 - 예) ~년 후/전, ~부터/까지/정도, 지난 세기
	DT_DURATION	- 기간 - 예) ~부터 ~까지, 일년 내내, ~ 간, 전반기/후반기, 성수기/비수기
	DT_DAY	- 날짜/절기 - 예) 입춘, 곡우, 첫날, 첫주, 내일, 모레, 금일, 당일
	DT_MONTH	- 달 - 예) 이달, 저번달, 올 4 월
	DT_YEAR	- 년 - 예) 작년, 내년, 올해, 서기 1910 년, 영락 1 년
	DT_SEASON	- 계절 - 예) 작년 봄, 올 가을, 매년 가을, ~시즌(예: 2008 년 시즌)
	DT_GEOAGE	- 지질시대 - 예) 원시시대, 구석기시대, 캄브리아기, 중생대, 원생대, 선사시대
	DT_DYNASTY	- 왕조시대 - 예) 청대, 조선시대, 조선 후기, 명대, 고려시대, 고려 말
TIME (TI)	TI_OTHERS	- 기타시간 - 예) 3 시이전

	TI_DURATION	- 기간 - 예) 6 시~9 시, 낮, 밤, 점심, 저녁, 오전, 오후, 밤낮, 전반전
	TI_HOUR	- 시각 - 예) 12 시, 저녁 8 시
	TI_MINUTE	- 분 - 예) 13 분
	TI_SECOND	- 초 - 예) 27 초
CIVILIZATION (CV)	CV_NAME	- 문명/문화 명칭 - 예) 인더스문명, 황허문명, 안데스문명, 거석문화, 도식문화
	CV_TRIBE	- 민족/종족 명칭, 국가를 구성하는 국민 - 예) 이라크족, 유대인, 조선족, 마오리족, 아메리카인디언, 고산족
	CV_SPORTS	- 스포츠/레포츠/레저 명칭 - 예) 축구, 야구, 수상스키, 스키, 스케이트, 하이킹, 행글라이딩
	CV_SPORTS_INST	- 스포츠 용품/도구 - 예) 스케이트, 인라인스케이트, 야구공, 베트, 아이젠, 오리발
	CV_POLICY	- 제도/정책 명칭 - 예) 카스트, 쓰레기종량제, 직업공무원제도, 연금제도, 고시제도
	CV_TAX	- 조세명칭 - 예) 법인세, 소득공제, 소득공제제도, 면허세, 관세, 부가가치세
	CV_FUNDS	- 연금, 기금, 자금, 펀드 명칭 - 예) 거치연금, 문예진흥기금, 공적자금, 퇴직연금, 코리아 펀드
	CV_LANGUAGE	- 언어명칭 - 예) 한국어, 아랍어, 일본어, 산스크리트
	CV_BUILDING_TYPE	- 건축양식 명칭 - 예) 기능주의건축, 노르만양식, 시토파건축, 도리스양식
	CV_FOOD	- 음식 명칭, 음식재료 - 예) 쇠고기, 소금, 어리굴젓, 명란젓, 국수, 현미, 김치, 나물
	CV_DRINK	- 음료수, 술 명칭 - 예) 위스키, 우유, 요구르트, 주스, 홍차, 레모네이드, 칵테일
	CV_CLOTHING	- 의복/섬유 명칭 - 예) 모시, 메리야스, 와이셔츠, 레이온, 양모, 점퍼
	CV_POSITION	- 직위/직책 명칭, 스포츠 포지션 - 예) 천황, 학사, 사장, 감독자, 장군, 왕, 교주
	CV_RELATION	- 인간관계명칭 - 예) 엄마, 아버지, 가족, 형제, 제자, 아들, 딸, 자식, 친척

	CV_OCCUPATION	-직업명칭 - 예) 화가, 극작가, 건축가, 가정교사, 기생, 언어학자, 공인중개사
	CV_CURRENCY	-통화명칭 - 예) 달러, 원, 엔, 유로, 루피, 위안
	CV_PRIZE	-상과훈장 - 예) 노벨경제학상, 대통령상, 대중상, 다승왕, 타격왕
	CV_LAW	- 법/법률 명칭 - 예) 국제연합헌장, 민법, 형법, 함무라비법전, 관습법, 헌법
	CV_FOOD_STYLE	-음식종류 - 예) 한식, 일식, 양식, 중식, 보양식, 패스트푸드, 육식, 채식
ANIMAL (AM)	AM_OTHERS	-기타동물명칭 - 예) 절지동물(예: 가재), 연체동물(예: 문어, 오징어)
	AM_INSECT	- 곤충 - 예) 벼메뚜기, 가시거미, 지네, 불나방, 배추흰나비, 참매미
	AM_BIRD	- 조류 - 예) 왜가리, 삿구기, 거위, 두루미, 황새
	AM_FISH	- 어류 - 예) 참가자미, 전기뱀장어, 연어, 갈치, 멸치, 미꾸리, 송어, 날치
	AM_MAMMALIA	- 포유류 - 예) 해달, 토끼, 오소리, 돌고래, 고래, 박쥐, 고슴도치, 두더쥐
	AM_AMPHIBIA	- 양서류 - 예) 개구리, 두꺼비, 산개구리, 산청개구리, 도롱뇽, 맹꽁이
	AM_REPTILIA	- 파충류 - 예) 도마뱀, 거북, 자라, 구렁이, 백사, 공룡, 악어
	AM_TYPE	-동물분류명칭 - 예) 강장동물, 구두동물, 극피동물, 내항동물, 모악동물, 선형동물
	AM_PART	- 동물 몸의 한 부분(신체부위) 명칭 - 예) 목, 살, 등, 배, 몸통, 몸, 손, 입, 갈기
PLANT (PT)	PT_OTHERS	-기타식물명칭 - 예) 우산이끼
	PT_FRUIT	- 과일 이름 - 예) 사과, 배, 자두, 바나나, 복숭아, 오렌지, 포도, 호두
	PT_FLOWER	-꽃이름 - 예) 저먼아이리스, 백합, 국화, 제비꽃, 히아신스, 초롱꽃, 할미꽃
	PT_TREE	-나무이름 - 예) 후박나무, 벚나무, 분비나무, 너도밤나무

	PT_GRASS	-풀이름 - 예) 향부자, 하늘타리뿌리, 인삼, 생강, 귀리, 감자, 상아, 갈대
	PT_TYPE	-식물유형명칭 - 예) 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
	PT_PART	- 식물의 한 부분에 대한 명칭 - 예) 뿌리, 가지, 가시, 결눈, 꽃, 열매, 잎, 줄기, 꽃가루, 꽃송이
QUANTIT Y (QT)	QT_OTHERS	- 기타 수량표현 - 예) 지식머니 20, 북위 38° 53' 51
	QT_AGE	- 나이 - 예) 만 55 세, 50~60 대
	QT_SIZE	- 크기/넓이 - 예) 9,606.33 km²
	QT_LENGTH	- 길이/거리/높이 - 예) 0.76mm, 900m, 8km, 82 피트
	QT_COUNT	- 개수/빈도 - 예) 6 개, 8 백 70 선, 하나, 둘
	QT_MAN_COUNT	- 인원수 - 예) 한명, 두사람, 3 명, 열댓명
	QT_WEIGHT	- 무게 - 예) 30 kg, 140g, 백근, 9 만 t, 85~100 kg, 60g, 5 만 톤, 170t
	QT_PERCENTAGE	- 백분율, 비율, 농도 - 예) 약 75%, 8~12ppm, 4 배, 400~500ppm, 1/3, 절반 이상
	QT_SPEED	- 속도 - 예) 1km/s, 30km/s
	QT_TEMPERATURE	- 온도 - 예) 22~25°C, 41~43.5°
	QT_VOLUME	- 부피 - 예) 두말, 2 홉, 2 되, 2 만 2000 km³, 20ℓ
	QT_ORDER	- 순서, 순차적 표현 - 예) 첫, 1 위, 재수/3 수
	QT_PRICE	- 금액 - 예) 72 만원, 60 억달러
	QT_PHONE	- 전화번호 - 예) 042-101-1010
	QT_SPORTS	- 스포츠 관련 수량표현 - 예) ~경기, ~승/패/무, 10 호골, ~안타

	QT_CHANNEL	- TV/라디오 채널번호 - 예) 2 채널, FM 98.1, 39 쇼핑, CJ 홈쇼핑
	QT_ALBUM	- 앨범관련수량표현 - 예) 1 집
	QT_ZIPCODE	- 우편번호 - 예) 110-704, 200-776
STUDY_FIEL D (FD)	FD_OTHERS	- 학문분야 및 학파 - 예) 농학, 군사학, 가정학, 건강과학, 교통학, 도서관학, 박물관학, 삼림학, 스포츠과학
	FD_SCIENCE	- 과학 학문분야 - 참고) 자연과학과 공학만으로 한정함 - 예) 자연과학: 물리학, 생명과학, 수학, 지구과학, 화학, 천문학
	FD_SOCIAL_SCIENCE	- 사회과학 학문분야 및 학파 - 정치/경제/사회와 관련된 분야 - 예) 학문분야: 인류학, 사회학, 경제학, 정치학, 심리학
	FD_MEDICINE	- 의학관련 학문분야 및 분과 - 학문분야: 의학, 병리학, 보건의료학, 신경과학, 가정의학 - 분과: 비뇨기과, 신경과, 가정의학과, 간호학과, 내과
	FD_ART	- 예술관련 학문분야 및 학파 - 예) 학문분야: 문예학, 음악사학, 미술사학, 공연예술학
	FD_PHILOSOPHY	- 철학관련 학문분야 및 학파 - 예) 학문분야: 교육철학, 형이상학, 윤리학, 정치철학, 과학 철학 - 예) 학파: 스킨라학파
THEORY (TR)	TR_OTHERS	기타 이론/법칙/원리
	TR_SCIENCE	- 과학 관련 이론/법칙/방식/양식/체계/학설 - 예) 21 세기형수학공식, 이진법, 2 진수, 3 원질설
	TR_SOCIAL_SCIENCE	- 사회과학 이론/법칙/방법/원리/사상, 정치사상 - 예) 민주주의, 공산주의, 노동가치설, 대화법, 법치주의, 평등주의
	TR_ART	- 예술관련 이론/법칙/방식/양식, 예술사조 - 예) 고딕양식, 고전주의
	TR_PHILOSOPHY	- 철학 이론/사상 - 예) 구조주의, 인식론, 존재론, 토마스주의, 경험주의, 합리주의
	TR_MEDICINE	- 의학 요법/처방, 의학 진단법 - 예) 인공호흡, 요혈자극법, 바이오피드백
EVENT (EV)	EV_OTHERS	- 기타 사건/사고 명칭, ~사태 - 예) 태안 기름유출사고, 북경세계대회

	EV_ACTIVITY	- 사회운동 및 선언 - 예) 공민권운동, 노예해방선언, 밀라노칙령, 인권선언
	EV_WAR_REVOLUTION	- 전쟁/혁명 - 예) 신해혁명, 포클랜드전쟁, 아편전쟁, 러일전쟁, 청교도혁명
	EV_SPORTS	- 스포츠/레저 관련 행사 - 예) 올림픽경기대회, 서울올림픽, 아시안게임
	EV_FESTIVAL	- 축제명칭 - 예) "~축제", 영화제만 대상으로 함
MATERIAL (MT)	MT_ELEMENT	- 원소명 - 예) 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨
	MT_METAL	- 금속물 - 예) 황산철, 알루미늄, 아연, 구리, 수은, 아말감
	MT_ROCK	- 암석명칭 - 예) 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암
	MT_CHEMICAL	- 모든 화학물질을 나타냄
TERM (TM)	TM_COLOR	- 색 - 예) 흰색, 흰빛, 흑색, 회색, 황, 홍색, 하늘색, 파란빛
	TM_DIRECTION	- 방향 - 예) 최북단, 중앙부, 중앙, 중부, 서쪽, 서부, 상부, 북쪽
	TM_CLIMATE	- 기후지역 명칭 - 예) 아열대 기후, 고산기후, 고지기후, 도시기후, 동안기후
	TM_SHAPE	- 모양/형태 명칭 - 예) 삼각형, 사각형, 팔각형
	TM_CELL_TISSUE	- 세포/조직/기관 명칭 (외부에서 안 보이는 조직/기관) - 예) 근육, 극세포, 간상세포, 딸세포, 신경세포, 표피, 정원세포
	TMM_DISEASE	- 증상/증세/질병 - 예) 기침, 폐결핵, 기관지확장증, 레트증후군, 피고름가래
	TMM_DRUG	- 약/약품명 - 예) 헤파린, 근이완제, 모기향, 소화제, 아스피린, 구충제
	TMI_HW	- IT 하드웨어 용어 - 예) 하드디스크드라이브, 마이크로프로세서, 롬, 디스켓
	TMI_SW	- IT 소프트웨어 용어 - 예) 트로이목마, 인터넷 웜, 웜바이러스, 님다바이러스
	TMI_SITE	- URL 주소 - 예) http://www.naver.com

TMI_EMAIL	- 이메일주소 - 예) forever@etri.re.kr
TMI_MODEL	- 각종 제품의 세부 모델명, 부품류 - 예) LM1000-2CX
TMI_SERVICE	- IT 서비스 용어 - 예) 와이브로 서비스, DMB 서비스, 위성 DMB, 카카오톡
TMI_PROJECT	- 프로젝트 명칭 - 예) BrainKorea21, 4 대강사업, 그린 IT 사업, 무료치과진료사업
TMIG_GENRE	-게임장르 - 예) RPG, 실시간 전략 게임, 1 인칭 슈팅 게임
TM_SPORTS	- 스포츠/레저 용어 (기술/규칙 명칭) - 예) 밀어내기 득점, 솔로 홈런, 바스켓카운트, 리바운드, 발리슛

본 사업에서는 위의 대분류 15 가지와 태그로 개체명을 태깅하고 개체명 분석 말뭉치를 구축하였다. 개체명 말뭉치를 구축할 때 주의해야 할 사항으로는 세 가지가 있다. 우선, 개체명은 무엇을 대상으로 하는가, 즉 개체명 태깅 대상에 대한 문제이다. 개체명(Named Entity)은 고유명사와 같이 이름 붙어진 실제 세계의 개체를 대상으로 한다. 따라서 개체명에는 모든 고유명사가 포함된다. 사전적으로 고유명사는 낱말의 특정한 사물이나 사람을 다른 것들과 구별하여 부르기 위해 붙인 고유의 이름을 의미하는 것으로, 실세계에 유일하게 존재하는 특정 ‘인스턴스(instance)’ 또는 ‘실체’의 명칭을 나타낸다. 그러나 개체명을 고유명사로 한정하면 인명, 지명, 작품명, 단체명 등 특정 범주에 대한 것으로 그 대상이 제한되어 버린다. 따라서 그 외의 범주는 고유명사가 아닌 세부 분류명이 개체명에 포함된다. 예를 들어 AM 과 PT 에는 동물과 식물의 다양한 종이 개체명 대상이 되며, MT 에는 원소나 금속, 암석의 구체적인 유형이, FD 에는 학문의 하위 분야가 개체명의 대상이 되는 등이다. 그러나 어떠한 경우이더라도 개체명 태깅 대상은 명사, 복합명사, 명사 구에 한정되며 다른 품사의 단어를 대상으로 하지는 않는다.

두 번째는 개체명 태깅 단위와 관련된 문제이다. 개체명은 단어뿐 아니라 ‘구’도 대상으로 하므로, 단어 단위로 태깅할 것인지 주변의 단어와 함께 구를 대상으로 태깅할 것인지가 문제가 된다. 일례로 “대전 유성구 신성동”의 경우, <대전 유성구 신성동/LC>과 같이 태깅할 수도 있고 <대전/LC> <유성구/LC> <신성동/LC>과 같이 태깅할 수도 있다. 이 경우 전체 구의 의미가 각 단어 의미의 합과 같다면, 즉 구의 의미가 투명하다면 최소 단위로 태깅하는 방식을 취한다. 따라서 본 지침에서는 두 번째 방식을 취한다. 다른 예로 “미국 컬럼비아 대학”의 경우 <미국/LC> <컬럼비아 대학/OG>으로 태깅하게 된다. 그러나 “한국 해양 대학”처럼 국가명이 고유명칭에 포함되어 있는 경우, ‘한국’과 ‘해양 대학’을 따로 태깅하면 본래의 의미를 갖지 못하므로 <한국 해양 대학/OG>과 같이 태깅한다.

개체명의 태깅 단위에서 한 가지 더 주의할 사항은 복합어(합성어·파생어) 내부의 구성요소는 개체명에 해당되더라도 따로 태깅하지 않는다는 점이다. 복합어는 그 자체로 구성요소와는 구별되는 별도의 개체를 지시하므로, 복합어 전체가 개체명일 경우에만 태깅하며 복합어가 개체명이 아닐 때에는 내부 구성요소가 개체명이더라도 별도로 태깅하지 않는다.

예를 들어 '야구장'의 '야구', '경찰차'의 '경찰'은 둘 다 CV에 속하나 합성어 내부 요소이므로 태깅하지 않는다. 이에 대한 일부 예외는 뒤에서 자세히 살필 것이다.

세 번째는 어떤 태그를 부여할 것인가, 즉 태깅 범주와 관련된 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. 본 지침은 중복 태깅을 허용하지 않으므로 문맥을 고려하여 가장 적절한 한 가지로 태깅을 해야 한다. 간혹 문맥으로 판정이 어려운 경우도 있으므로 이러한 유형은 따로 목록화하여 우선순위를 배당할 것이다. 본 지침에서는 15 가지 개체명 범주에 대해 하나씩 살피면서 상기의 세 문제점들에 대해 정리할 것이다. 지침의 마지막에서는 뒤의 두 문제, 즉 개체명 태깅 단위 및 중의적이어서 우선순위를 결정해 주어야 하는 태깅 범주에 대한 문제를 다시 모아 정리·제시할 것이다.

나. 세부분류 개체명 정의

1. PERSON

PERSON 은 실존인물, 신화속 인물, 게임/소설의 가상 캐릭터 등을 나타낸다.

1.1 PS_NAME

- 정 의
 - 인명/인명별칭
 - ◆ 실존 인물
 - ◆ 신화/종교 등장인물
 - ◆ 게임/소설 등의 가상 캐릭터 인물
- 실존인물
 - 예 제
 - ◆ 흥선대원군, 해원 신윤복, 현종, 프랭클린 루스벨트, 콘스탄티누스, 콘스탄티누스 1세, W.가이거, S.A.키르케고르, F.M.도스토예프스키, 마르쿠스 아우렐리우스
 - ◆ 가수 인물: 소녀시대, 아이유
 - ◆ 배우 인물: 원빈, 한가인, 장동건
 - ◆ 감독 인물: 봉준호, 박찬욱
 - 본 문
 - ◆ 또한 쿤은 S.A.키르케고르나 M.하이데거의 실존론적(實存論的) 사고방식에 따라 미적 기분을 다루었다.
 - ◆ 한편 키르케고르는 데모니셰를 죄악감에 빠진 인간이 악(惡)에서 헤어날 수 없는 채 선(善)에 대하여 품는 불안감이라고 설명하였다.
 - ◆ 1802 년 독일 유학에서 돌아온, H.스테펜스가 독일 낭만주의 정신을 고취하게 되자 A.월렌슐레거를 선두로 낭만주의 운동이 일어나, N.그룬트비 ·안데르센 ·키르케고르 등이 이어받았다.
 - ◆ 그의 철학은 S.키르케고르와 K.야스퍼스 계열에 속하는 그리스도교적 실존주의였다.
- 신화/종교 등장인물
 - 예 제
 - ◆ 티에스테스, 보현보살
 - ◆ 아스페르스, 신농, 엔키두, 웅녀, 전륜성왕, 아가멤논

■ 본 문

- ◆ 티에스테스는 미케네의 왕이 된 아트레우스의 아내 아이로페를 유혹하였고, 아트레우스는 그를 추방하였다가 나중에 속임수를 써 다시 불러들인 다음, 그의 세 아들을 죽여 요리로 만들어 먹게 하였다.
- ◆ 보현보살이 타고 있는 동물로 등장하기도 하고
- ◆ 아스페르스는 오랫동안 조국을 갖지 못한 유대민족의 상징으로 '영원한 유대인'이라는 별명까지 생겼다.
- ◆ 신농씨는 화덕(火德)을 가지고 있었기 때문에 염제(炎帝)라 하였는데,
- ◆ 엔키두는 길가메시에게 안긴 채 죽었다.
- ◆ 곰은 참아내, 삼칠일(三七日 즉 21 일) 만에 여인으로 변하여 웅녀가 되었다.
- ◆ 전통적으로 마우리아왕조의 아소카왕[阿育王](BC 3 세기)을 세속의 전륜성왕이라고도 말한다.

- 게임/소설 등의 가상 캐릭터 인물

■ 예 제

- ◆ 짱구, 심슨, 슈퍼맨, 배트맨

- 주 의

- '성+씨'의 경우 특정한 사람을 지시할 때 '성'만 PS_NAME 으로 태깅함. 가문을 의미할 때에는 태깅하지 않음.
 - ◆ 왕 씨 → 왕/PS_NAME
- 지위 + 인명의 경우, 지위와 인명을 따로 인식함
 - ◆ "국대비 왕씨"의 경우, <국대비:CV_POSITOIN>, <왕:PS_NAME>으로 인식함.
 - ◆ "폐비 윤씨"의 경우, <폐비:CV_POSITOIN>, <윤:PS_NAME>으로 인식함.
- 별명이나 별칭이 특정한 인물을 지시하는 경우, 그 별명과 별칭이 인명 카테고리 안에 해당될 경우만 PS 로 태깅한다.
 - ◆ '베이브 류스' 류현진 완벽 투구 → 베이브류스/PS
 - ◆ 고교생 신데렐라/PS 의 이미지를 벗고 → 신데렐라/PS
 - ◆ '베이브 류스', '신데렐라'의 경우 별명, 별칭이어도 원래 'PS'에 속하기 때문에 'PS'로 태깅한다.
- 별명이나 별칭이 특정한 인물을 지시하더라도 그 자체로는 인명 카테고리에 속하지 않는 경우, 본래 단어가 속한 개체명 분류에 맞춰서 태깅한다.
 - ◆ 양파/PT 총리/CV, 나무늘보/AM (지침 비유적 표현 123 쪽 참고)

2. STUDY_FIELD

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

학문 분야는 위키백과의 분류체계를 따라 다음과 같이 나눈다.

위키 대분류	위키 세분류	개체명 세부분류
인문학	언어학	FD_OTHERS
	역사학	FD_SOCIAL_SCIENCE
	종교학	
	철학	FD_PHILOSOPHY
	문예학	FD_ART
	음악사학	
	미술사학	
	공연예술학	
사회과학	정치학	FD_SOCIAL_SCIENCE
	경제학	
	사회학	
	행정학	
	고고학	
	인류학	
	지리학	
	심리학	
	문헌정보학	
자연과학	물리학	FD_SCIENCE
	생명과학	
	우주과학	
	지구과학	
	화학	
응용과학	경영학	FD_SOCIAL_SCIENCE

	교육학	
	신학	
	언론학	
	법학	
	의학	FD_MEDICINE
	공학	FD_SCIENCE
	농학	FD_OTHERS
	군사학	
	가정학	
	건강과학	
	교통학	
	도서관학	
	박물관학	
	삼림학	
	스포츠과학	

2.1 FD_SCIENCE

- 정 의

- 과학 관련 학문 분야 및 과학 학파
- 과학분야는 크게 순수과학과 기술과학으로 분류된다. 순수과학이란 엄밀한 뜻의 자연과학을 응용과학에 대응해서 쓰는 말이다. 순수과학의 종류로는 물리학, 생명과학, 수학, 지구과학, 화학 등이 있다. 이에 반해 기술과학이란 일종의 응용과학을 말하는 것으로 건축공학, 금속재료공학, 기계공학, 무기 및 병기공학, 산업공학, 수의학, 원자력공학, 자동차공학, 전기공학, 전자공학, 조선해양공학, 토목건설공학, 항공우주공학, 화학공학, 환경공학 등이 이에 해당한다.

- 예 제

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

- 본 문

- 생물학은 생물의 종류 ·구조 ·기능, 그 밖의 연구대상에 따라 여러 가지 분야로 나눌 수 있지만,
- 유전학은 생물의 형질 발현에 대한 원인을 구명하는 과학으로
- 응용지질학은 옛날에는 주로 광물자원의 생성과정과 성인(成因) 및 경제성과 분포에 관해 연구하는 광상학(鑛床學:economic geology)으로 대표되었다.
- 고등학교 교육과정의 지구과학은 이와 같은 넓은 뜻의 지구과학을 가리킨다.

2.2 FD_SOCIAL_SCIENCE

- 정 의

- 사회과학 관련 학문 분야 및 학파
- 사회과학이란 인간 사회의 여러 현상을 과학적, 체계적으로 연구하는 모든 경험과 학을 말한다. 사회과학의 분야로는 경영, 경제, 교육, 국방 및 군사, 뉴스와 미디어, 법, 사회, 정치, 통계, 행정 등이 포함된다.

- 예 제

- 경영학, 국제경영학, 신고전학파
- 사회인류학, 기능학파, 뒤르캄학파, 통계학
- 인류학, 사회학, 경제학, 정치학, 심리학, 인문지리학, 법학, 응용사회과학, 경영학, 고학, 행정학, 문헌정보학, 역사학, 언어학, 종교학, 교육학, 신학, 언론학, 가정학

- 본 문

- A.마셜을 창시자로 A.C.피구, J.M.케인스, D.H.로버트슨, J.V.로빈슨 등을 후계자로 하여 케임브리지대학교를 중심으로 발달하였기 때문에 케임브리지학파(Cambridge school)라고도 한다. 케임브리지학파가 신고전학파로 불리는 것은, 기본적으로는 이 학파가 고전학파의 해체나 한계혁명(限界革命) 후의 영국의 정통파였으며, 한계효용 가치론을 채용하면서도 다른 학파에 비하여 공급면의 분석이나 장기동태면의 관심 등에서 고전학파의 전통을 가장 많이 승계하고 있기 때문이다
- 경제학은 이와 같은 실천적 측면을 가지고 있는 것이다.
- 오늘날 사회인류학(social anthropology)이라는 학문 분야는 기능학파의 영향을 받은 것이다.
- 프랑스의 뒤르캄학파는 집단표상(集團表象)을 중시하므로 경제현상을 사회학적 방법으로 파악하였고,
- 통계학은 19 세기 후반부터 20 세기 초에 걸쳐

2.3 FD_MEDICINE

- 정 의
 - 의학 관련 학문 분야 및 학파
 - 의학이란 인간을 질병으로부터 구하고 건강법을 모색하는 학문이다.
 - 의학 학문분야는 양의학과 한의학을 포함하여 각 학문별 세부 학문 명칭을 나타낸다.
- 예 제
 - 학문분야
 - ◆ 의학, 병리학, 보건의료학, 신경과학, 가정의학, 감각생리학, 기초의학, 생리학, 해부학, 생리체학, 기생충학, 면역학, 미생물학, 병리학, 생화학, 약리학, 의료정보학, 의사학, 법의학, 치의학, 약학, 한의학, 기상의학, 구강위생학
 - ◆ 보건학, 약학, 내분비학, 병리학, 열대의학
 - ◆ ~의학과
 - 분과
 - ◆ 비뇨기과, 신경과, 가정의학과, 간호학과, 내과, 소아청소년과, 정신과, 피부과, 재활의학과, 안과, 이비인후과, 외과, 산부인과, 영상의학과, 핵의학과, 방사선종양학과
 - ◆ ~외과
- 본 문
 - 오늘날과 같이 구강위생학으로서 자리를 굳히기 시작한 것은 제 2 차 세계대전 후부터의 일이다.
 - 보건학은 환경위생, 전염병 관리, 개인위생 교육, 질병의 조기진단과 예방을 위한 의료서비스조직,
 - 약학은 약을 중심으로 하여 화학 ·물리학 ·생물학 등의 분야를 포함하며, 그 연구수단에 따라 학문적 체계를 구성하고 있다.
 - 오늘날까지 이 개념이 내분비학의 기초가 되고 있다
 - 병리학에서는 현실적으로 병리해부학과 실험병리학에 중점을 두고 있다.
 - 열대병을 대상으로 하는 의학이 바로 열대 의학이라고 하겠다.
- 주 의
 - 비뇨기 등은 FD_MEDICINE 이 아님
 - 의료 학과(분과)의 경우에는 FD_MEDICINE 으로 태깅한다.
 - ◆ 예) 그렇다면 정형외과와 재활의학과를 체험 하고..

2.4 FD_ART

- 정 의

- 예술 관련 학문분야 및 학파(유파)
- FD_ART 는 예술 관련 학문 분야 외에도 예술의 분류, 장르가 태깅 대상에 포함이 된다. 무용, 미술, 연극, 영화, 문학, 사진, 음악, 현대예술의 각 세부 예술 형태의 분류 명칭이 FD_ART 에 해당한다. 미술의 경우 공예, 디자인, 서예, 조각, 판화, 회화 등이 이에 해당한다.
- 각 세부 예술 형태의 분류 명칭이 FD_ART 에 포함된다고 해도 예술의 '장르'로 나왔을 때만 한정해서 태깅하며, 예술 활동의 '결과물(작품)'은 태깅하지 않는다.

- 예 제

- 학문분야
 - ◆ 문예학, 음악사학, 미술사학, 공연예술학
- 학파
 - ◆ 인상파, 점묘파, 우키요에 화파, 신경향파
- 예술 분류
 - ◆ 현대음악, 현악 4 중주, 동양화, 서양화, 한국화, 수채화, 조각, 공포영화, 과학영화
 - ◆ 소설, 뮤지컬, 음악, 영화, 연극, 미술, 팝송, 시
 - ◆ 음악 장르: 아악, 군악, 힙합, 범패, 가곡, 판소리
 - ◆ 회화 장르: 소묘, 크로키
 - ◆ 소설 장르: 우화, 판타지소설, SF 소설

- 본 문

- 일본의 춘화는 우키요에 화파의 등장과 더불어 활발하게 제작되었는데
- 신경향파에 드는 주요 작품을 들면 다음과 같다.
- 한국어로 보통 시라고 할 때에는 그 형식적 측면을 주로 가리켜 문학의 한 장르로서 시작품(poem)을 말하는 경우가 있다.

- 주 의

- '노래', '춤'은 일반적인 명사로 FD_ART 대상이 아님.
 - ◆ 세부적인 장르인 '뮤지컬', '발레' 등이 FD_ART 의 대상임
 - ◆ 춤 유형, 춤의 종류도 FD_ART 로 태깅한다. 예) 재즈댄스/FD, 검무/FD, 사교댄스/FD
- '신화'도 FD_ART 로 태깅함.
- 예술의 '장르'가 아닌 예술 활동의 '결과물(작품)'은 태깅하지 않는다.
 - ◆ '사진 700 장을 발견했다', '김영하 작가의 신작 소설'에서 '사진, 소설'은 예술 활동의 결과물이기 때문에 태깅하지 않는다.
 - ◆ '가격보다 디자인이 예쁜 상품이 좋다'에서 '디자인'도 예술 장르라고 보기 어렵기 때문에 태깅하지 않는다.

2.5 FD_PHILOSOPHY

- 정 의
 - 철학 관련 학문분야 및 학파(유파)
 - 철학이란 인생, 세계 등등에 대해 연구하는 학문으로 크게 논리학, 동양철학, 서양철학, 심리학, 윤리학 등으로 분류된다.
- 예 제
 - 학문분야
 - ◆ 교육철학, 형이상학, 윤리학, 정치철학, 과학철학, 언어철학, 논리학, 예술철학, 미학, 윤리학, 사회철학, 스콜라철학, 주기철학, 과학철학, 성리학, 심리학
 - 학파
 - ◆ 스콜라학파, 플라톤 학파, 신프로이트파, 유가, 도가, 묵가, 법가, 명가, 음양가, 종횡가, 병가
- 본 문
 - 아테네학파의 최후를 장식하는 플라톤학파는 의식적으로 엄밀한 정의·공준(公準)·공리에서 출발하여 논증을 하는 해석법을 연구하여 그 후의 수학 발전에 크게 공헌하였다.
 - 심리학이 연구대상으로 삼는 '마음'이라는 것은
 - 신프로이트파의 일반적인 특징은 리비도를 생물적·성적으로 보는 프로이트의 고정설을 거부하고,
 - 윤리학은 일반적으로 인간의 행위에 관한 여러 가지 문제와 규범을 연구하는 학문이라고 할 수 있다.
- 주 의
 - '빈 학단'과 같은 철학 단체는 FD_PHILOSOPHY 대상이 아님
 - "~ 주의"의 경우에는 이론(THEORY) 분류에 해당되고, "~주의 학파"의 경우에 학문 분야(FIELD) 분류에 해당됨
 -
 -

2.6 FD_OTHERS

- 정 의
 - 기타 학문분야 및 학파
 - FD_SCIENCE, FD_SOCIAL_SCIENCE, FD_MEDICINE, FD_ART, FD_PHILOSOPHY의 세부 학문분야 이외의 학문분야 및 학파
- 예 제
 - 농학, 군사학, 가정학, 건강과학, 교통학, 도서관학, 박물관학, 산림학, 스포츠과학
- 본 문
 - 대장금 번역한 대만인 동문군 씨, 국어국문학 박사학위 취득.
 - 인문학의 위기 이후로 독어독문학, 불어불문학에 대한 관심이 줄어들고 있다.

3. THEORY

THEORY 는 특정 이론, 법칙, 원리 등의 명칭을 의미한다. THEORY 는 명확히 기술될 수 있는 고유의 이론, 법칙, 원리만을 말하며, 일반적인 용어들(예: 사회법칙, 법이론 등)은 해당하지 않는다. 세부 분류로는 TR_SCIENCE, TR_SOCIAL_SCIENCE, TR_ART, TR_PHILOSOPHY, TR_MEDICINE 으로 분류된다. 이 세부 분류를 벗어나는 이론, 법칙 명칭은 TR_OTHERS 에 포함된다.

- 예 제

- 점진주의, 피마준법, 존왕양이 사상, 봉건적 배외사상, 존왕론, 양이론, 천왕숭배 사상

- 주 의

- THEORY 는 '~ 사상, ~ 원리, ~ 이론'의 형태로 기술될 수 있다. 이 경우 '~ 사상, ~ 원리, ~ 이론'을 포함해서 THEORY 로 간주한다.
- THEORY 는 '천황숭배' 사상의 형태로 기호와 함께 쓰일 수 있다. 이 경우 기호를 포함해서 THEORY 로 간주한다.

3.1 TR_SCIENCE

- 정 의

- 순수과학: 과학 관련 이론/법칙/방식/양식/체계/학설
- 기술과학: 기계/장치/시스템/기술 관련 이론/법칙/방식
- 단순한 작용/현상은 TR_SCIENCE 의 대상으로 보지 않음
 - ◆ 단순 작용/현상으로 TR_SCIENCE 의 대상이 아닌 예제
 - 형광작용, 중력, 빙식작용, 관성력
 - X 선은 사진작용·형광작용·이온화작용, 진공 상태에서는 빛과 같은 속도로 진행한다는 성질이 있고, 그 외에도 입자와 같이 회절한다.
 - 중력의 크기는 물체의 질량에 비례하며, 질량 1 g 의 물체에는 대체로 980 dyn 의 중력이, xg 의 물체에는 980 dyn 의 x 배의 중력이 미친다. 보통 물체의 무게라고 하는 것은 이 힘을 가리킨다.
 - 빙식작용은 크레바스(베르크슈르트라고도 한다)의 밑바닥에서 융빙수(融氷水)의 동결에 의해 행하여지는 파쇄작용과, 빙하의 유동에 따라 이루어지는 연마작용의 두 가지 방법으로 기반암석(基盤岩石)을 침식한다.
 - 그러므로 관성력은 가속도 운동을 하고 있는 관측자에게만 느껴진다. 관성력은 힘의 원인인 물체가 없는 가상의 힘이므로 작용·반작용의 법칙이 성립하지 않는다.
 - ◆ 특정 인물이 발견한 고유한 현상으로, 인물 이름이 붙은 현상/작용은 TR_SCIENCE 대상으로 태깅함 (예: 도플러 효과)

- 순수과학
 - 예 제
 - ◆ 특수상대성이론, 우열의 법칙, 양자론, 멘델법칙, 드무아브르의 정리, 상대성원리
 - ◆ 21 세기형수학공식, 이진법, 3 원질설
 - 본 문
 - ◆ 아인슈타인은 이 특수상대성이론을 발전시켜 1916 년 일반상대성이론을 발표하였다.
 - ◆ 운동량이나 좌표를 연산자로서 생각하는 양자론과 대비된다.
 - ◆ 동의유전자도 멘델법칙의 변형으로 설명할 수 있다.
 - ◆ 특수상대성이론은 운동에 관한 갈릴레이-뉴턴의 상대성원리를 근본적으로 개혁하여, 서로 등속도로 운동하는 관측자에 대하여 전자기파의 이론을 포함한 모든 물리법칙이 같은 형식으로 기술되도록 정식화(定式化)되어 있다.
 - 주 의
 - ◆ '2 진수', '자연수', '약수' 등은 개체명 대상이 아님
- 기술과학
 - 예 제
 - ◆ 원심분리법, 기체확산법, 열확산법, 컨베이어 라인(conveyer line) 시스템
 - ◆ 진공건조, 이온주입법, 공랭식, 수랭식
 - 본 문
 - ◆ 농축 우라늄의 제조법으로는 기체확산법 ·열확산법 ·원심분리법 등의 여러 방법이 이용된다.
 - ◆ 1913 년 포드자동차에서 처음 적용한 컨베이어 라인(conveyer line) 시스템이 최초이다.
 - ◆ 일반적으로 행하는 진공건조는 총체적으로 이와 같은 현상을 이용하고 있으며, 원료 중의 수분이 많은 상태에서 비교적 저온으로 건조시킬 수 있다.
 - ◆ 이온주입법은 이 불순물반도체를 만드는 방법 중 하나이다.
 - ◆ 냉각 방식에는 공랭식과 수랭식이 있는데
 - 주 의
 - 고유의 명칭을 갖고 있는 특정 과학 이론만이 태깅 대상이며, 여러 구체적인 이론을 포함할 수 있는 포괄적인 개념의 이론/원리는 태깅 대상이 아니다. 다음은 부적절한 TR_SCIENCE 의 예제들이다.
 - ◆ 서로 등속도로 운동하는 관측자에 대하여 전자기파의 이론을 포함한 모든 물리법칙이 같은 형식으로 기술되도록 정식화되어 있다.
 - ◆ 현재는 양자역학과 더불어 현대물리학, 특히 소립자물리학이나 우주론의 지도적 원리로 간주되고 있다.
 - 이론이 아니라, 특정 용어나 개념에 해당되는 것은 TR_SCIENCE 가 아님
 - ◆ 예) 자기장, 적외선

3.2 TR_SOCIAL_SCIENCE

- 정 의
 - 사회과학 이론/법칙/방법/원리/사상, 정치사상
- 예 제
 - 민주주의, 공산주의, 다원적 민주주의, 법치주의, 평등주의, 관료주의, 민족주의, 자본주의, 사회주의, 개인주의, 자유주의
 - 무작위추출법, 노동가치설, 대화법
 - 하인리히의 법칙, 수요와 공급의 법칙
 - 농업방식: 삼포식 농업
 - 북학파, 신고전학파, 고전학파
- 본 문
 - 개방된 시장경제와 다원적 민주주의라는 가치관을 공유하는 국가간 경제사회 정책 협의체로서
 - 그러나 실제로는 많은 문제가 그대로 남아 있어 미국 민주주의의 문제점이 되고 있다.
 - 오늘날 공산주의라고 할 때는 문헌에만 남아 있는 죽은 공산주의가 아니라,
 - 이 무작위추출법을 이용하여 조사대상 전체를 조사하는 대신,
 - 노동가치설은 공산주의자들에 의해 지지(支持)되고 있다.
 - 대화법의 원류는 고대 소크라테스의 교육방법에서 찾을 수 있으며,
 - 사회주의란 말이 개인주의의 반대말로서 르루에 의하여 만들어졌음을 알 수 있다.
- 주 의
 - 개체명 태깅 기본 원칙에 따라 최장 유형을 태깅 대상으로 한다. 아래의 예에서 '신고전학파 성장론'이 이론 명칭이 된다. 여기서 '신고전학파'라는 FD_SOCIAL_SCIENCE 가 포함되어 있으나 최장 개체명 태깅 원칙에 따라 '신고전학파 성장론'을 TR_SOCIAL_SCIENCE 로 태깅한다.
 - ◆ 본문 예제: 신고전학파 성장론의 경우에는 생산요소의 대체성(가변성)이 특히 주목되고 있으며, P.A.새뮤얼슨이 주장하여 영향을 미친 신고전파종합 (neoclassical synthesis)의 경우에는 경제적 복지의 달성수단으로서의 사기업제와 경쟁의 유효성에 초점이 맞추어져 있다고 생각된다.
 - '사회주의', '공산주의' 등의 정치사상도 TR_SOCIAL_SCIENCE 에 해당됨
 - '개인주의'도 자유주의 정치사상의 하위범주이므로 TR_SOCIAL_SCIENCE 에 해당됨
 - '로리타 콤플렉스', '오이디푸스 콤플렉스' 등의 콤플렉스는 심리학의 이론이므로 TR_SOCIAL_SCIENCE 에 해당됨

3.3 TR_ART

- 정 의
 - 예술 관련 이론/법칙/방식/양식, 예술사조
- 예 제
 - 에로티시즘, 원근법, 대기원근법, 선원근법, 구름법, 물골법
 - 대위법, 데카당스, 문자기보법, 환유법, 제유법
 - 고딕양식, 고전주의, 유량악보, 낭만주의
 - 몽타주, 오마주, 추사체, 데칼코마니, 액션페인팅
- 본 문
 - 원근법은 넓은 의미로 볼 때 화면 깊이의 표현이며, 대기원근법은 선원근법의 보조적 효과를 가질 뿐만 아니라, 화면에 중심과 통일을 주어 고전양식을 성립하게 한다.
 - 쌍구라고도 한다. 이에 대하여 선, 즉 골을 나타내지 않고 그리는 방법을 물골법이라고 한다.
 - 영화에서 에로티시즘은 존재의 문제에 접근하려는 시도를 보였으나 때로는 성행위를 노골적으로 묘사한 포르노그래피에 불과한 경우도 많았다.
 - 대위법에서는 각 성부가 명료하게 식별할 수 있는 선율적 독립성을 지니며,
 - 이후 그들의 예술적 경향을 데카당스라고 평하였다.
 - 이 문자기보법에 의한 악보를 문자보라고 한다.
 - 환유법은 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.
- 주 의
 - “다다이즘” 등의 예술운동은 일종의 예술사조로 고려해서 TR_ART 로 태깅함
 - “르네상스”는 예술운동 외의 학문, 사상 등까지 다루고 있으므로 예술사조로 보기 힘들기 때문에 사회 운동으로 고려해서 “EV_ACTIVITY”로 태깅함
-
-

3.4 TR_PHILOSOPHY

- 정 의
 - 철학 이론/사상
- 예 제
 - 조작주의, 주지주의, 개인주의, 구조주의, 토마스주의, 경험주의, 합리주의, 공리주의, 현실주의
 - 니힐리즘, 이데아설, 인식론, 노장사상, 이기론, 존재론, 관념론
- 본 문
 - 엄밀한 의미에서의 니힐리즘은 아무것도 존재하지 않는다는, 즉 무라는 주장이다.
 - 조작주의는 원래 물리학의 상대성이론에서 여러 개념을 규정하는 데 사용되었으나,
 - 이데아설을 소크라테스의 것이라고 하여 버넷-테일러설을 세웠다.

- J.로크의 《인간오성론》은 이 인식문제의 전환점에 선 것이었고, D.흄 등의 영국 경험론에 의하여 인식론의 근대적 성격은 더욱 명확하게 되었으며 I.칸트에서 대성되었다.
- 그러나 그들의 사상이 도가사상을 실질적으로 대표하므로 노장사상을 일반적으로 도가사상과 같은 것으로 보기도 한다.
- 이에 따라 이와 기를 유기적으로 결합한 이기론이 확립되었다
- 인식이 감각에 의한 것이 아니라 지성에 의해서 생긴다고 보는 합리론도 넓은 뜻의 주지주의이다.

- 주 의

- “플라톤 학파” 등의 철학 유파는 FD_PHILOSOPHY 이다.
- ‘유가, 도가, 묵가, 법가, 명가, 음양가, 종횡가, 병가’ 역시 학파이기 때문에 FD 로 태깅함.

3.5 TR_MEDICINE

- 정 의

- 의학 요법/처방, 의학 진단법

- 예 제

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

- 본 문

- 수지침의 방법으로는 기초기능조절 상응요법·기맥요법·보사법·오치처방·삼일체질진단·기맥연결요법·팔성혈 자세교정법 등의 방법 등이 있다
- 345 개의 치료점에서 중요한 것은 기맥의 유주를 따른 보사법이 있고 요혈자극법이 있다
- 이 발열요법은 매독 스피로헤타가 고열에 약하다는 사실에 착안하여 발상된 것으로
- 간단한 견인요법으로는 반창고의 응용이 있으며,
- 보법의 운용은 주로 체질과 기능의 부족을 보충하여 일체의 쇠약현상을 제거하는 법이다
- 대기요법은 결핵환자가 자연환경에서 맑은 대기를 가까이하는 생활을 통하여 치료하는 방법이고, 공기요법은 일정한 시간에 피부를 드러내어 공기를 접하게 하는 방법이다.
- 이외에 염파요법, 수지력 테스트에 의한 진단과 치료 및 전자진찰도 있으나 더욱 중요한 것은 음양맥진이다.
- 촉각으로 환자의 질병을 판단하는 방법인 절진 중 맥진을 가장 중요시하기 때문에 절진은 곧 맥진을 말한다.
- 내시경검사 영역은 주로 기관지·식도·위·복강·흉강·요도·방광·대장·직장·자궁·관절 등이며,
- 혈액검사도 이 범주에 속하는 중요한 검사이다.

- 이 중에서 검안의 주체가 되는 것은 굴절도의 검사인데,
- 문진의 내용은 형식적으로는 기왕력·가족병력·주소·현병력 등으로 나눈다.

3.6 TR_OTHERS

- 정 의
 - 기타 이론/법칙/원리
 - TR_SCIENCE, TR_SOCIAL_SCIENCE, TR_ART, TR_PHILOSOPHY, TR_MEDICINE 의 세부 이론 분야 이외의 이론/법칙/원리
- 예 제
 - 샤머니즘, 원칙주의
- 본 문
 - 토착종교인 샤머니즘의 고원교를 몰아내는 투쟁과 그 승리를 극명하게 표현한다.

4. ARTIFACTS

ARTIFACTS 는 인공물을 나타낸다. 인공물이란 사람에 의해 창조된 대상을 말하며 문화재, 건물, 악기, 도로, 무기, 운송수단, 작품명 등이 이에 해당한다. 유의할 점은 ARTIFACTS 내부에 포함된 세부범주에 해당되지 않는 인공물은 태깅 대상에서 배제한다는 점이다.

4.1 AF_CULTURAL_ASSET

- 정 의
 - 문화재 명칭
- 예 제
 - 경복궁, 보현사, 경주 서악리 마애석불상 등
 - 불국사, 석굴암, 관상감 측우대, 문양전
 - 무령왕릉, 진흥왕릉, 무열왕릉, 천마총, 서봉총
- 본 문
 - 관상감 측우대는 이 측우기를 받치기 위하여 만든 것으로 추정되며, 지금은 대석만 남아 있다.
 - 불국사 경내에는 다보탑(多寶塔, 국보 20), 3 층석탑(국보 21), 연화교·칠보교(蓮華橋·七寶橋, 국보 22), 청운교·백운교(靑雲橋·白雲橋, 국보 23), 금동비로자나불좌상(金銅毘盧遮那佛坐像, 국보 26), 금동아미타여래좌상(金銅阿彌陀如來坐像, 국보 27), 사리탑(舍利塔, 보물 61) 기타 많은 문화재가 잔존한다.
 - 현세(現世)의 부모를 위하여 불국사를 세우는 한편, 전세(前世)의 부모를 위해서는 석굴암을 세웠다는 것이다.
 - 이 일련의 문양전들은 모두 환원염(還元焰)으로 구워낸 회색 부전(敷塼)으로, 비교적 견지(堅緻)한 바탕흙[胎土]이 사용되었으며, 신라의 것과는 달리 한면에 문양을 양각(陽刻)하였다.
- 주 의
 - 문화재는 다양한 형태를 지닌다. 역사적인 건물, 서적, 작품 등이 모두 문화재에 속할 수 있다. 따라서 AF_CULTURAL_ASSET 대부분이 ARTIFACTS 의 다른 세부 개체명에 속하게 된다. 이 경우 문화재보다는 다른 세부 개체명을 우선시하여 태깅한다. 그러나 AF_BUILDING 의 경우 문화재에 해당된다면 AF_CULTURAL_ASSET 로 태깅한다. 따라서 사찰, 고궁 등은 모두 문화재로 간주한다.
 - 문화재가 LC 등 다른 범주의 개체 유형으로 할당 가능한 경우, 역시 다른 개체명으로 태깅한다. 다만, LC_OTHERS 등의 OTHERS 계열이나 세부 개체유형이 없는 경우에만 AF_CULTURAL_ASSET 로 태깅한다.
 - “광주칠석고싸움놀이”, “강강술래”와 같은 무형문화재도 세부 개체유형이 있으면 세부 개체유형으로 하고, 세부 개체유형이 없는 경우에만 AF_CULTURAL_ASSET 로 태깅함
 - 전통의식/무속의식/제사 등은 AF_CULTURAL_ASSET 이외의 세부 개체유형으로 태깅 가능하면 세부 개체유형으로 태깅함. 문화재로 지정된 전통의식/무속의식/제사 중에서 세부 개체유형이 없는 경우에만 AF_CULTURAL_ASSET 로 태깅함

- 자연유산 중 LC_TOUR, PLANT 계열, ANIMAL 계열 등의 세부 개체유형으로 되는 것들은 세부 개체명으로 태깅함. 천연기념물의 경우에도 세부 개체유형이 있으면 LC_TOUR, PLANT 계열 등의 세부 개체유형으로 태깅하고, 세부 개체유형이 없거나 LC_OTHERS 와 같이 OTHERS 계열인 천연기념물만 AF_CULTURAL_ASSET 로 태깅함
- 문화재로 지정되어 있지는 않지만 조선 시대 이전에 만들어진 절, 무덤은 AF_CULTURAL_ASSET 으로 태깅함 (예: 서봉총, 부석사, 황룡사)
- "이형 좌명원종공신녹권 및 함"처럼 서책과 함(기타 유물)이 함께 있어서 1 개의 세부 개체유형으로 특정할 수 없는 문화재는 전체를 AF_CULTURAL_ASSET 으로 태깅함
- 유네스코 세계기록유산은 유네스코와 세계기록유산을 구분하여 태깅
 - ◆ <유네스코:OG_OTHERS> <세계기록유산:AF_CULTURAL_ASSET>
- 보물 1 호, 국보 1 호의 경우는 AF+QT 로 나눠서 분석하지 않고 통합하여 태깅한다.
 - ◆ 보물 1426 호: 보물 1426 호/AF

4.2 AF_BUILDING

- 정 의
 - 건축물/토목건설물, 운동장이름, 아파트, 다리, 등대, 분수
- 예 제
 - 대한생명 63 빌딩, 구대한의원 본관 등
 - 서해대교, 대청댐, 충주댐, 롯데월드타워, 동대문운동장
 - 뽕순감옥
- 본 문
 - 서해대교의 사장교는 국내 최대 규모이며,
 - 소양강댐·안동댐·대청댐·충주댐 등 대부분은 다목적댐들이다.
 - 동대문운동장과 장충체육관 등 기존시설만으로는 국내외의 스포츠 수요에 대응하기 어려움을 감안하여 서울시가 1976 년 10 월 건설기본계획을 수립하였다.
 - 법원은 롯데월드타워를 임시 거처로 하되 공사가 끝나면 다시 돌아가라고 결정했고 신 명예회장은 6 월 롯데호텔로 복귀했다.
- 주 의
 - 도로 명칭은 AF_ROAD 에 해당된다. 빌딩, 다리, 댐 명칭 등이 AF_BUILDING 에 해당한다.
 - 문화재인 절은 AF_CULTURAL_ASSET 에 포함된다.
 - '뽕순감옥'은 AF_BUILDING 이고, '뽕순감옥 형장'은 LC_OTHERS 로 태깅한다.
 - 어떠한 단어가 건축물이면서 기관의 의미를 지닐 때, 문맥에 따라 건축물의 개념으로 사용되었다고 판단되는 경우, OG 가 아닌 AF 로 태깅한다.
 - '수영장'이나 '체조경기장' 같은 건축물이 지명이나 상위 조직과 함께 쓰인 경우, 전체가 정식 명칭이라면 분리하지 않고 하나의 덩어리로 태깅한다.
 - ◆ '한강수영장'은 수영장 정식 명칭 자체가 '한강수영장'이기 때문에 분리해서 태깅하지 않는다.
 - ◆ 예) 한강 잠원지구 한강수영장 : 한강/LC +**잠원지구/LC** +한강수영장/AF

4.3 AF_MUSICAL_INSTRUMENT

- 정 의
 - 악기 명칭
- 예 제
 - 기타, 캐스터네츠 등
 - 편경, 팽과리, 대금, 해금, 파이프오르간, 오보에
- 본 문
 - 중국의 석경보다 좋은 편경을 만들어 내게 되었다.
 - 팽과리가 종묘제례악(宗廟祭禮樂)에 쓰일 때는 소금(小金)이라는 다른 이름으로 불린다.
 - 이 위에 갈대 속청을 붙여 이것으로 대금 특유의 음색을 낸다.
 - 해금은 현악기이면서도 관악합주에 반드시 편성되어 관악기와 현악기와의 균형을 유지하며
 - 파이프오르간은 보통 케이스(상자)에 의해 분리된 몇 개의 파이프군(群)이 복합되어 하나의 악기를 이루고 있다.
 - 오보에란 이름은 프랑스어의 '음이 높은 나무피리'라는 뜻의 오브와(hautbois)에서 나왔다.

4.4 AF_ROAD

- 정 의
 - 도로/철로 명칭
 - 운하
 - 거리
- 예 제
 - 시베리아철도, 경부고속도로, 강원선, 호남선, 경원선, 경의선 등
 - 세종로, 올림픽대로, 충북선
 - 일리노이-미시시피운하, 글리비체 운하
 - 거리: 명동거리, 샹젤리제거리
- 본 문
 - 세종의 시호를 따서 세종로라는 명칭을 붙였다.
 - 도로 이름을 올림픽대로로 고쳤다.
 - 충북선의 수송수요가 증대함에 따라 이의 복선화(113 km 구간) 계획이 추진되어
- 주 의
 - 한국에서 2014 년부터 정식 시행된 새로운 주소 표기법인 도로명주소법에는 도로명이 주소로 표시된다. 도로명주소법에서의 도로명은 AF_ROAD 로 태깅함
 - 주소에서 도로명 다음에 나오는 번지는 QT_OTHERS 로 독립적으로 태깅함 (예: 남대문로 39 → <남대문로:AF_ROAD> <39:QT_OTHERS>)
 - 같은 곳을 나타내는 주소여도 지번과 도로명 주소를 구분하여 지침대로 태깅한다.
예) 국립국어원 주소

- ① 서울특별시/LC+ 강서구/LC+ 방화동/LC+ 827/QT
- ② 서울특별시/LC+ 강서구/LC+ 금남화로/AF+ 154/QT
- '11 번 국도, 44 번 국도'의 경우 고유의 번호를 가진 국도이기 때문에 통합하여 태깅한다.
예) 11 번 국도 : 11 번 국도/AF
- 제주 해안도로, 강릉 해안도로, 우회도로, 산간도로, 산악도로, 해안도로의 경우 특정한 것을 지칭하지 않기 때문에 태깅하지 않는다.
- '남산 둘레길', '제주 올레길'의 경우 특정 명칭을 포함하고 있기 때문에 통합하여 태깅한다. (단, 둘레길, 올레길이 단독으로 나왔을 경우에는 태깅하지 않는다.)
예) 남산 둘레길 : 남산 둘레길/AF

4.5 AF_WEAPON

- 정 의
 - 무기 명칭
- 예 제
 - 기관단총, 시한폭탄, 곡사포, 고정포 등
 - 승자총통, 탄도미사일, 최루가스, 스마트폭탄
 - 무기 유형(일반명사) 포함: 도끼, 칼, 전투기
- 본 문
 - 1583 년(선조 16) 6 월 승자총통을 처음 만들었다는 기록이 있다.
 - 탄도미사일이 최초로 실전에 사용된 것은 제 2 차 세계대전 중에 독일이 개발하여 런던 공격에 사용된 V-Ⅱ호였다.
 - 최루가스를 사용할 경우에는 가스탄 ·투척 ·투사기(投射器) 등이 사용된다.
 - 스마트폭탄은 베트남전쟁에서 처음으로 사용되어 효용이 증명되었다.
- 주 의
 - 특정 무기 명칭과 무기 유형 명칭 모두 해당된다.
 - 최루 가스는 화학 물질이긴 하나, 무기로 쓰이는 것이 일반적이기 때문에 무기로 태깅하는 것이 맞음
 - 문맥에 따라 무기 유형을 볼 수 있는 일반명사도 태깅할 수 있다. 문맥상에서 무기로 사용되었다고 판단될 경우 AF 로 태깅할 수 있음.

4.6 AF_TRANSPORT

- 정 의
 - 교통수단/자동차/선박 모델 및 유형
 - 운송수단 및 놀이기구
- 예 제
 - 갈라테아호, 그레이트브리튼호, 경비정, 가스터빈자동차, 메르세데스벤츠
 - 자기부상열차, 나룻배, 경자동차, 탈리스

- 교통수단 유형(일반명사) 포함: 자동차, 배, 비행기, 선박
- 본 문
 - 자기부상열차는 1994 년 현재 독일과 일본을 중심으로 개발되고 있는데,
 - 큰 강이나 사람이 직접 건너지 못하는 깊은 냇물에는 대개 나룻배가 있었다.
 - 일반적으로 경자동차라고 하면 4 륜의 승용차를 말한다.
 - 1997 년 12 월 파리와 브뤼셀 사이를 잇는 새로운 노선을 개통하면서 탈리스로 명명하였다.
- 주의 사항
 - 우주탐사선, 우주선은 AF_TRANSPORT 이고, 인공위성은 TMI_HW 로 태깅함
 - ◆ 과학기술 3 호 → TMI_HW
 - ◆ 마리너 6 호, 갈릴레오호, 메신저호 → AF_TRANSPORT
 - ◆ 컬럼비아호, 챌린저호, 디스커버리호 AF_TRANSPORT
 - TMI_MODEL 은 카메라 모델명(D800) 등의 특정 모델명으로 우주선, 인공위성은 대상이 아님
 - '장보고함' 등의 함선도 AF_WEAPON 이 아니라 AF_TRANSPORT 로 태깅함

4.7 AF_WORKS

- 정 의
 - 기타 작품명
- 주의
 - '조선일보'와 같은 신문사는 OGG_MEDIA 에 해당됨.
 - 세부 인공물 분류에 포함되지 않는 작품명의 경우 AF_WORKS 에 해당한다.

4.7.1 AFW_DOCUMENT

- 정 의
 - 도서/서적 작품명
 - 지리서, 지도
 - 의학관련 서적
 - 종교서적, 각 종교의 경전
 - 철학관련 서적
 - 시/소설/희곡 등의 문학 작품명
 - 역사서적 (예: 중종실록, 고려사)
 - 기타 서적 (예: 뽀뽀한 영철영어)
 - 기타 문서 (예: 미국 독립 선언서)

- 지리서, 지도

■ 예 제

- ◆ 신증동국여지승람, 대동여지도, 각도여지도, 동국여지승람 등
- ◆ 아방강역고, 택리지, 혼일강리역대국도지도, 곤여만국전도

■ 본 문

- ◆ 《신증동국여지승람》에 전라도의 토산물로 기록되어 있다
- ◆ 1864 년(고종 원년)에 재간된 《대동여지도》 수정본으로 거창박물관 2 층 중앙에 전시되어 있다.
- ◆ 옛 기록에 '대부는 전라도 함열의 대위향(大位鄕)이라'고 하였으며, 《동국여지승람(東國輿地勝覽)》에 '대위향은 함열 남쪽 6 리 지점에 있다'고 하였는데, 지금의 익산시 함열읍 지역 내에 있었을 것으로 짐작된다.
- ◆ 다산(茶山) 정약용(丁若鏞)이 저술한 《아방강역고(我邦疆域考)》를 1903 년 위암(韋庵) 장지연(張志淵)이 증보한 것으로,
- ◆ 《택리지》라는 이름도 후인들이 그 내용을 보고 붙인 이름인 듯하다.
- ◆ 한국 지도 중에는 이회(李薈) 등이 1402 년에 제작한 《혼일강리역대국도지도(混一疆理歷代國都地圖)》가 있다.
- ◆ 원래의 곤여만국전도는 1602 년 이탈리아의 예수회 선교사 마테오 리치가 중국에서 명(明)나라 학자 이지조(李之操)와 함께 목판에 새겨 인쇄한 목판본이다.

- 의학 서적

■ 예 제

- ◆ 동의보감, 의림비서
- ◆ 본초강목, 의방신감, 삼화자향약방, 의약감

■ 본 문

- ◆ 《자산어보(玆山魚譜)》에 보면, 맛이 달콤하고 회·국·포를 만들기 좋다고 했으며, 《동의보감(東醫寶鑑)》에 성(性)이 평(平)하고 맛이 달며 독이 없다고 했다
- ◆ 명나라의 이시진(李時珍)이 쓴 《본초강목(本草綱目)》에는 인체 각 부위의 약효까지 자세히 설명되어 있다.
- ◆ 《동의보감》 《의종손익》 《방약합편》 등 30 여 권의 동의학책들을 참고하고 자신의 임상경험을 기초로 《의방신감》을 편찬했다.
- ◆ 이것을 편찬하는 데 기본이 된 의서가 《삼화자향약방》이라는 사실을 밝히고 있다.
- ◆ 《각세신편팔감》은 11 권 6 책으로 되어 있는데, 마지막 11 권이 이 의약감이다.

- 종교 서적

■ 예 제

- ◆ 한역대장경, 장아함경, 중아함경, 잡아함경, 증일아함경, 흠정역성서, 개역성서 등
- ◆ 코란, 요한의 묵시록, 아베스타

■ 본 문

- ◆ 이 이름으로 불려지는 문헌이 현재 《한역대장경(漢譯大藏經)》 안에 네 가지가 포함되어 있는데 ① 《장아함경(長阿含經)》 ② 《중아함경(中阿含經)》 ③ 《잡아함경(雜阿含經)》 ④ 《증일아함경(增一阿含經)》이다
- ◆ 그의 가르침은 《아함경》《율장》 등의 원시불교 경전을 통해 전하여지고 있다
- ◆ 이슬람교도(무슬림)는 일상생활에서 《코란》을 독송하며, 각종 의식에서는 정규 독송자가 소리 높이 독송한다.
- ◆ 그리스도교의 대표적인 계시문학서가 바로 이 《요한의 묵시록》이다.
- ◆ 페르시아 고전인 조로아스터교의 경전 《아베스타》에 나타나 있다.

- 철학 서적

■ 예 제

- ◆ 존재와 시간, 노자도덕경 등
- ◆ 범주론, 명제론, 토정비결, 장자, 예기

■ 본 문

- ◆ 《노자(老子)》 또는 《도덕경(道德經)》이라고도 한다.
- ◆ 이런 의미에서 도와 덕을 설파하는 데서, 《노자》의 가르침은 도덕(오늘날의 도덕과는 다름)으로 불리어 《도덕경》이라는 별명이 생기게 되었다
- ◆ 아리스토텔레스의 저작들 중에서 논리학서로 분류되는 것은《범주론》《명제론》《분석론 전서》《분석론 후서》《토피카(장소론)》《귀변론》《소피스트 논박》 등이 있다.
- ◆ 정월 초승이면 으레 《토정비결》로 그 해 신수를 알아보는 일은, 조선 민간의 세시풍경이다.
- ◆ 그의 심현한 철학사상서이자 우수한 문학서인 이 《장자》는 위(魏)·진(晉) 때에 널리 읽히고 육조시대(六朝時代)까지 그 사상이 유행하였다.
- ◆ 《주례(周禮)》《의례(儀禮)》와 함께 삼례(三禮)라고 하며 《의례》가 예의 경문(經文)이라면 《예기》는 그 설명서에 해당한다.

- 시/소설/희곡 등의 문학 작품명

■ 예 제

- ◆ 돈카를로스, 쿠오바디스, 구스타프 3 세 등
- ◆ 반지의 제왕, 난장이가 쏘아올린 작은 공, 모란이 피기까지는

■ 본 문

- ◆ 대본은 프랑스의 극작가 A.E.스크리브의 《구스타브 3 세》를 A.소마가 이탈리아어로 번역한 것이다
- ◆ 원작인《반지의 제왕》은 1954 년 출간된 이래 1 억 권 이상이 팔린 초베스트셀러로,
- ◆ 장편 《난장이가 쏘아올린 작은 공》은 12 편의 작품으로 구성된 연작소설로 도시빈민의 가난한 삶과 처참한 패배의 한을 다루고 있다.

- ◆ 이어 《내 마음 아실 이》《가늘한 내음》《모란이 피기까지는》 등의 서정시를 계속 발표하였고, 1935년에는 첫째 시집인 《영랑시집(永郎詩集)》을 간행하였다.

- 역사 서적
 - 중종실록, 고려사
- 기타 서적
 - 뽀뽀한 영철영어
- 주 의
 - 만화, 만화책은 AFW_DOCUMENT 에 해당하고, 만화 애니메이션은 AFW_VIDEO 에 해당함
 - "연행가, "일동장유가", "상춘곡" 등의 가사는 AFW_DOCUMENT 로 태깅함
 - 서적으로서의 <훈민정음>은 AFW_DOCUMENT 로 태깅하지만, 문자적인 의미의 '훈민정음'은 개체명이 아니므로 태깅하지 않음.
 - '로미오와 줄리엣', '맥베스', '오텔로' 등의 희극/비극 등은 문학작품으로 AFW_DOCUMENT 로도 볼 수 있고, 연극으로 AFW_PERFORMANCE 로도 볼 수 있다. 명확히 연극작품의 작품명으로 볼 수 있는 경우에만 AFW_PERFORMANCE 로 태깅하고, 그 이외에는 AFW_DOCUMENT 로 태깅함

4.7.2 AFW_PERFORMANCE

- 정 의
 - 춤/무용 작품명 (예: 백조의 호수)
 - 춤 종류는 AF 로 태깅하지 않고 FD_ART 로 태깅한다. AF_PERFORMANCE 에서 춤/무용은 구체적인 작품명만 해당됨
 - 연극/공연명/가극 (예: 캐츠, 돈카를로스, 피바다)
 - 뮤지컬
- 춤/무용 작품명
 - 예 제
 - ◆ 농악무, 동래학춤, 바라춤, 간척무 등
백조의 호수, 세에라자드, 파소도블레
 - 본 문
 - ◆ 파리에서《사육제》《불새》와 더불어《세에라자드》를 공연하여 성공을 거두었다.
 - ◆ 파소도블레는 스페인의 전통 춤의 일종이다. 남녀가 짝을 지어 8분의 6 박자라는 빠른 템포에 맞춰 1분에 약 120번 스텝을 밟는다.

■ 주 의

- ◆ 특정 춤/무용 작품명만 AF 로 태깅한다. 춤 종류는 FD_ART 로 태깅한다. 따라서 '재즈댄스, 검무, 사교댄스, 에어로빅댄스' 등은 FD 로 태깅한다.

- 연극/공연명

■ 예 제

- ◆ 돈카를로스, 피바다, 밀림아 이야기하라 등
- ◆ 명성황후, 캐츠, 마이 페어 레이디

■ 본 문

- ◆ 북한이 내세우는 5 대 혁명가극, 즉 《피바다》 《꽃파는 처녀》 《당의 참된 딸》 《밀림아 이야기하라》 《금강산의 노래》 의 하나로 《피바다》 를 연출했던 김영희가 연출했다
- ◆ 처음으로 인민상을 받은 작품은 1960 년에 창작된 가극 《밀림아 이야기하라》 이며
- ◆ 《여우사냥》 은 작품 전체를 노래화한다는 연출자의 의도에 따라 61 곡의 노래로 짜여진 《명성황후》 로 재창조되었다.
- ◆ 1990 년 내한한 영국산 뮤지컬 《캐츠》 는 예술의 전당 오페라극장에서 서구 뮤지컬의 진수를 만끽하게 해주었다.
- ◆ 또한 이 극은 뮤지컬 《마이 페어 레이디》 의 원작이다.

- 주 의

- 뮤지컬은 연극 형식에 음악이 추가된 것이므로 AFW_PERFORMANCE 로 태깅하고, 오페라는 음악 형식에 춤이 추가된 것이므로 AFW_MUSIC 으로 태깅함

4.7.3 AFW_VIDEO

- 정 의

- 영화 작품명
- TV 프로그램 이름 (예: 아빠 어디가, 9 시 뉴스, 1 박 2 일)
- 애니메이션 작품명
- TV, 라디오 채널명

- 세부 예제: 영화 작품명

■ 예 제

- ◆ 사랑방 손님과 어머니, 다이하드, 쉬리, 박하사탕 등
- ◆ 바람의 검심, 올랜도, 마리아 이야기, 취화선

■ 본 문

- ◆ 제 1 회에서는 신상옥(申相玉) 감독이 연출한 《연산군》 이 작품상을 수상했으며, 역시 신상옥 감독이 연출한 《사랑방 손님과 어머니》 가 감독상·시나리오상·특별장려상 등을 수상했다.
- ◆ 원제는 《바람의 검심-메이지 검객 낭만담 ろうに劒心-明治剣客浪漫譚》 이다.

- ◆ 셸리 포터는 《올랜도》발표 이후 국제적인 명성을 얻었으며, 1960년대 이후 페미니즘 영화의 큰 지평을 열었다.
- ◆ 《마리아 이야기》는 공간감과 온기를 살리기 위해 3D로 구조를 잡은 배경을 일일이 2D로 리터치하고, 그 위에 캐릭터와 특수효과를 합성하는 방식으로 작업했다.
- ◆ 영화의 제목은 세속적인 삶을 초월한 천재화가를 '술에 취해 그림을 그리는 신선'이라고 표현하여 취화선이라고 붙인 것이다.
- 세부 예제: TV 프로그램명
 - 예제
 - ◆ 아빠어디가, 9시뉴스, 1박2일
 - 본문
 - ◆ (주)엠플레이는 자사의 온라인게임 '크레이지 아케이드'를 온게임넷 "크아 한쌍, 크아 한판"이란 프로그램을 통해 방송한다고 밝혔다.
 - ◆ '생방송! 팡팡 게임짱'을 통해 방영되는 '통스통스'는 개그맨 정용국과 게임자키 레이가 공동으로 총 24회분량을 방영할 예정이다.
- 세부 예제: TV, 라디오 채널명
 - 예제
 - ◆ CJ 홈쇼핑, 코디미 TV, KBS 2TV, MBC sports+
 - 본문
 - ◆ 류현진 이번 시즌 경기 역시 MBC sports+에서 볼 수 있다.
 - 4일 오후 7시 씨티극장에서 채널 CGV를 통해 초청된 시청자를 포함한 관객들을 대상으로 '친절한 금자씨' 영화상영회를 갖고 박찬욱 감독과의 대화 시간도 마련한다.

4.7.4 AFW_ART_CRAFT

- 정 의
 - 미술 작품명(그림/미술품)
 - 조각품 등 조형 작품
 - 불교에서 나오는 그림명, 불상명
- 예 제
 - 생각하는 사람, 도산도, 도산서원도, 풍류절창, 춘투도, 춘화사계첩, 원앙비보 등
 - 키스, 게르니카, 지옥의 문
- 본 문
 - 1606년에 간행된 《풍류절창(風流絶暢)》은 24개의 채색판화로 구성된 것으로서 당인의 《춘투도(春鬪圖)》를 모본으로 제작한 것이라고 서문에 밝히고 있다
 - 그 사례로서 단원 김홍도(檀園 金弘道)의 서명과 낙관이 찍힌 《춘화사계첩(春畵四季帖)》을 들 수 있다.
 - 성문화가 개방적이었던 명대에 제작된 춘화첩 중 가장 뛰어난 채색목판화집으로 1642년에 간행된 《원앙비보(鴛鴦秘譜)》를 손꼽을 수 있다
 - 그의 다른 작품의 포옹하는 여인들과 비교할 때 《키스》가 긍정적인 인간의 성취를 표현하고 있다는 해석에는 반론의 여지가 없는 것 같다.

- 피카소는 이 조국의 비보에 접하자, 한 달 반 만에 대벽화를 완성, '게르니카'라고 이름붙였다.

- 《지옥의 문》의 청동 작품은 모두 로댕의 사후에 제작된 것이다.

- 주 의

- "풍속화첩" 등의 그림을 모은 화첩은 AFW_DOCUMENT 가 아니라 AFW_ART_CRAFT 에 해당됨
- "고려산적석사비(高麗山積石寺碑)", "광개토대왕릉비" 등의 비(碑)도 미술품으로 고려해서 AFW_ART_CRAFT 로 태깅함 (현재 태깅말뭉치에는 문화재로 태깅된 경우가 많으므로 수정 필요함)
- 비(碑)에 새겨진 비문도 AFW_ART_CRAFT 로 태깅함

4.7.5 AFW_MUSIC

- 정 의

- 음악 작품명

- 주 의

- 오페라 작품도 해당
- 고려속요도 해당됨

- 예 제

- 운명의 힘, 돈 카를로스 등
- 아이다, 카르멘, 한국환상곡, 보태평, 정대업

- 본 문

- 《운명의 힘》(1862) 《돈 카를로스》(1867) 등과 더불어 1859 년에 제작된 베르디의 중기 작품에 속한다
- 주세페 베르디(Giuseppe Verdi)의 대표적 오페라 <아이다(Aida)>는 1871 년 카이로에 세워진 이탈리아 극장 개관 기념으로 초연됐다.
- 오페라 코미크로서의 《카르멘》 등은 노래 아닌 대사가 포함되어 있지만 오페라에 포함시키고 있다.
- 1957 년 귀국하여 자작곡 《강천성악(降天聖樂)》 《한국환상곡》 등을 지휘하고
- 종묘제례악으로 채택된 《보태평》과 《정대업》은 500 여 년 동안 전승되면서 변화는 있었지만 현재까지 연주되고 있다.

- 주의 사항

- "도술가", "서동요" 등의 향가는 AFW_MUSIC 으로 태깅함
- 뮤직 비디오 작품도 AFW_VIDEO 가 아니라 AFW_MUSIC 에 해당됨

4.8 AF_WARES

- 정 의
 - 상품/제품 이름
- 예 제
 - 운동화 이름, 화장품 이름, 여행상품 이름, 게임 상품 이름 (구체적인 상품 이름만 태깅함)
- 본 문
 - 우리은행도 하이-믹스(Hi-Mix) 복합예금 32 호를 500 억원 한도로 다음달 3 일까지 판매한다.
 - 이 기간, TNGT 매장에서 10 만원 이상 구매 고객에게 그레이드 '크리스탈로맨스'를 증정한다.
 - '리그 오브 레전드'가 10 월 온라인게임 점유율 45%를 넘었다.
- 주 의
 - '아이폰, 갤럭시에스'와 같은 구체적인 하드웨어 상품명은 개념적으로 AF_WARES 에 속하기도 하고 TM_HARDWARE 에 속하기도 함. 따라서 정확한 상품명으로 나온 경우는 AF 로 태깅하고 그 외 일반 명사에 해당하는 제품들은 TM 으로 태깅하는 것으로 구분함.
 - ◆ '아이패드, 아이폰, 갤럭시에스'는 구체적인 상품명이므로 모두 AF 로 태깅함.
 - ◆ 스마트폰, TV, 사운드바의 경우 하드웨어의 종류를 나타내는 일반 명사이므로 TM 으로 태깅함.
 - 티셔츠 '렛츠고 투게더'가 월드컵 개막 전에 40 만장이나 : 렛츠고 투게더/AF
 - IT 관련 상품도 TM 이 아닌 상품명에 해당하는 AF 로 태깅한다. 다른 개체명 태그셋과 겹칠 경우, 다른 개체명 태그셋을 우선 함.
 - TMI_HW, CV_FOOD, CV_DRINK, TMI_MODEL 등으로 태깅 가능하면 그쪽으로 태깅
 - 화장품이나 향수 관련 제품 등은 특별한 태그셋이 없으므로 이 태그를 사용

5. ORGANIZATION

ORGANIZATION 은 기관/단체 명칭을 나타내며, 경제기관, 교육기관, 군사기관, 미디어/방송 기관, 스포츠기관, 예술기관, 의료기관, 종교기관, 과학기관, 도서관, 법률기관, 정치/정부기관, 음식 관련 기관, 숙박기관 등으로 세분화된다. 세부 기관에 포함되지 않는 기관/단체들은 OG_OTHERS 로 태깅된다.

- 주 의

- 학파는 STUDY_FIELD 로 분류된다.
- 특정 기관/단체명과 그 기관/단체의 세부 조직 명칭이 동시에 쓰였을 경우 이를 묶어 전체를 기관/단체명으로 보지 않고 대표 기관/단체명만을 태깅한다.
 - ◆ “서울대학 응용미술학과 졸업” 에서 서울대학 응용미술학과는 기관/단체명으로 볼 수 없다. “서울대학”만을 기관/단체명으로 한다.
 - ◆ “공정거래위원회 사무처”는 ‘공정거래위원회’까지만 기관명으로 볼 수 있다.
- 조약, 협정, 협약은 OGG 계열의 관련 개체명으로 태깅한다. (대부분은 OGG_POLITICS 로 태깅함)
- 단, 협상은 EV 로 태깅한다.

5.1 OGG_ECONOMY

- 정 의

- 경제 관련 기관/단체
- 회사/기업
 - ◆ 은행, 카드사, 보험사, 증권사
 - ◆ 사기업: 삼성그룹, 삼성전자, LG 그룹, 현대자동차 등

- 주 의

- “다보스 포럼” 등의 정기적으로 개최되는 회의이면서 세부 조직을 가지고 있는 경우에는 EV_OTHERS 가 아니라 OGG 계열로 태깅함 (예: 다보스 포럼/OGG_ECONOMY)

- 경제 기관/단체

- 예 제
 - ◆ 공정거래위원회, 금융감독원, OECD, 국제통화기금, 석유수출국기구
- 본 문
 - ◆ 1994 년 12 월 공정거래위원회 및 공정거래위원회 사무처가 경제기획원으로부터 독립해 중앙행정기관으로 승격해 재정경제부 소속기관이 되었다가, 다시 국무총리 소속기관으로 바뀌었다.
 - ◆ 금융감독원에 대하여 금융기관에 대한 검사를 요청할 수 있도록 하고 있다.
 - ◆ 경제협력개발기구조약(OECD 조약)에 서명함으로써 OECD가 탄생하였다.
 - ◆ 국제통화기금은 세계무역의 안정된 확대를 통하여 가맹국들의 고용증대, 소득 증가, 생산자원개발에 기여하는 것을 궁극적인 목적으로 한다.

- ◆ 석유수출국기구(OPEC)는 생산량이 많고 표준품질인 아라비안 라이트(Arabian light)를 기준원유로 해왔다.
- 주 의
 - ◆ 기관/단체명은 공식적인 명칭에 한한다. 하위 조직을 포함해서 기관/단체명으로 하지 않는다.
- 회사/기업
 - 예 제
 - ◆ BYC(주), 포드자동차 등
 - ◆ 한국도로공사, 네슬레, 홀렛팩커드, 동양척식주식회사
 - ◆ 한국은행, 비씨카드, LG 화재, 동양증권
 - 본 문
 - ◆ 1946 년 한흥메리야스로 출발하여 1955 년 한흥산업(주)으로 상호를 바꾸었고 1957 년 백양상표를 등록하였다. 1960 년 한흥물산(주)으로 상호를 변경하였으며, 1975 년 기업을 공개하였다. 1977 년 올마크를 획득하였고, 1979 년 상호를 (주) 백양으로 변경하였다가, 1996 년 다시 (주) 비와이씨로 변경하였다.
 - ◆ 현재 한국도로공사가 이를 위임받아 도로의 유지관리를 전담하고 있는데, 모두 가 포장도로이다.
 - ◆ 1877 년 앵글로스위스는 네슬레의 유아식품을 생산하기 시작하였고, 네슬레도 앵글로스위스의 연유를 생산, 두 회사는 치열한 경쟁에 돌입하였다.
 - ◆ 홀렛팩커드(Hewlett-Packard Co. : HP)는 컴퓨터, 인터넷·인트라넷솔루션, 서비스, 통신제품 및 측정솔루션 등 여러 첨단정보사업 분야에서 탁월한 제품의 성능과 지원 면에서 세계적인 명성을 얻고 있는 초우량 글로벌기업이다.
 - ◆ 동양척식주식회사는 1908 년에 일본이 조선의 토지와 자원을 수탈할 목적으로 설치한 식민지(植民地) 착취기관이다
 - 주 의
 - ◆ 구로공단은 LC_OTHERS 에 해당됨
 - ◆ 경제 관련 공공기관/정부기관은 OGG_POLITICS 에 해당됨
 - 예) 근로복지공단, 한국산업단지공단
 - ◆ “빈폴”과 같은 브랜드도 OGG_ECONOMY 에 해당됨
 - ◆ 방위산업체는 특정한 기업을 지칭하는 것이 아니기 때문에 개체명 대상이 아님

5.2 OGG_EDUCATION

- 정 의
 - 교육 기관/단체
 - 학교, 학원, 어린이집, 훈련소, 학회
- 예 제
 - 홍익대학교, 국제연합교육과학문화기구, 라이프치히 음악원
 - 국자감, 대한지리학회, 국군간호사관학교, 한인중앙학원
 - 고용훈련소, 논산훈련소

- 본 문

- 1949 년 홍익대학으로 개편, 1961 년 대학정비령에 따라 일시 홍익미술대학으로 개칭하였다가 이듬해 홍익대학으로 환원하였으며
- 경상남도 진주에서 태어나 1950 년 경남중학을 졸업하고, 1954 년 부산대학교 법학과를 졸업하였으며, 1964 년에는 서울대학교 행정대학원에서 석사학위를 취득했다. 1984 년 일본 와세다대학 대학원에서 도시행정학을 연수하였고, 1991 년 계명대학교 대학원에서 법학박사 학위를 받았다.
- 교육기관이던 경학(京學)을 국자감으로 개칭하여 설치하였다.
- 창립 당시에는 조선지리학회라고 하였으나 1949 년 11 월 총회에서 대한지리학회로 바꾸었다.
- 1980 년 1 월 국군간호사관학교로 명칭을 변경하고 간호전문대학 과정으로 승격하였다.
- 1913 년 이승만이 교장으로 부임하면서 한인중앙학원으로 이름을 바꾸었다.

- 주 의

- '유네스코'와 같이 교육, 문화, 과학의 여러 분야와 관련된 기관은 OGG_OTHERS 로 태깅함
- 학교 학과의 경우에는 OGG_EDUCATION 으로 태깅하지 않음. 개체명 대상이 아님
 - ◆ 예) 1954 년 <부산대학교:OGG_EDUCATION> 법학과를 졸업하였으며....
 - ◆ 학과는 학교에 포함이 되어 있는 개념이기 때문에 학과는 제외
 - ◆ 캠퍼스 역시 동일.
 - 한양대학교 서울캠퍼스에서는 서울캠퍼스는 개체명 대상이 아님. 한양대학교만 개체명으로 태깅.
 - 홍익대 미대의 경우 미대는 개체명 대상이 아님. 홍익대만 태깅
 - ◆ 다만 ~ 도서관의 경우 앞에 사람 이름이 나오고 도서관이 나온 경우 사람 이름만 태깅을 하면 안됨. 사람이름과 도서관은 개체명 태그상 완전히 다른 의미이며 도서관을 뺐을 때 ~도서관이라는 개체명의 의미가 사라짐. 반면, 부산대학교 법학과는 법학과를 빼도 부산대학교라는 큰 의미는 변화가 없음. 앞의 단어와 뒤의 단어가 포함관계인지를 고려하는 것이 필요.
- ~ 훈련소는 모두 OGG_EDUCATION 으로 태깅한다. '고용훈련소'와 같이 특정 이름이 부여되지 않은 경우라도 OG 로 태깅한다.
- 학회의 경우에는 학회의 종류에 OGG_EDUCAION 이외의 세부적인 태그가 가능한 경우에는 OGG 세부 분야로 태깅하고, 세부 태그가 없는 경우에는 OGG_EDUCATION 으로 태깅한다.
 - ◆ 예) 한국정보과학회: OGG_SCIENCE
- '웨스트포인트사관학교' 등의 군사학교도 OGG_EDUCATION 으로 태깅함
- '광주특수전교육단' 등의 군사 교육을 시키는 기관도 OGG_EDUCATION 으로 태깅함

5.3 OGG_MILITARY

- 정 의
 - 군사 기관/단체 및 유형, 테러조직/무장단체
 - 군대 유형(일반명사)도 포함됨: 해군, 공군, 육군, 인민군
- 예 제
 - 아르헨티나군, 대한민국재향군인회 등
 - 국방품질관리소, 한미연합사령부, 국방홍보원, 국제연합군
- 본 문
 - 5 월 20 일 유엔 사무총장의 조정교섭이 실패로 돌아가자 영국군은 포클랜드에 상륙, 75 일간의 격전 끝에 6 월 14 일 아르헨티나군의 항복으로 전쟁을 종결시켰다
 - 1965 년 6 월 8 일 제 49 차 국무회의에서는 대한민국재향군인회가 WVF 에 가입한 5 월 8 일을 '재향군인의 날'로 정할 것을 의결하였다.
 - 1999 년 1 월 다시 국방품질관리소로 명칭을 바꾸었다.
 - 양국의 합의로 설치된 한미군사위원회의 전략지시에 의해 한미연합사령부가 창설되었다.
 - 2000 년 8 월 18 일 국방홍보원으로 명칭이 바뀌었다.
 - 첫째는 집단안전보장의 강제행동을 임무로 하는 국제연합군이며, 둘째는 국제연합의 평화유지활동의 일환으로 현지에 파견되는 새로운 형태의 국제연합군이다.
- 주 의
 - '탈레반', '이슬람국가' 등의 무장단체도 OGG_MILITARY 에 해당됨
 - 독립운동 단체 중에 전투 단체는 OGG_MILITARY 에 해당됨 (예: 오적암살단, 대한독립군)

5.4 OGG_MEDIA

- 정 의
 - 미디어 기관/단체
 - 방송관련 회사/기업
- 예 제
 - 국제잡지협회, 국제기자연맹, KBS, MBC, TJB 등
 - 방송심의위원회, 공익광고협의회, 스카이라이프(주)
- 본 문
 - 1980 년까지는 방송심의위원회의 업무를 한국방송윤리위원회가 수행하였다.
 - 1988 년 발전적으로 해체하고 공익광고의 자율운영계획을 수립하여 현재의 명칭인 공익광고협의회로 재구성하였다
 - 스카이라이프(주)는 2000 년 12 월 19 일 위성방송 사업권을 획득한 이후 2001 년 3 월부터 본격적인 위성방송 준비를 시작하여

- 주 의

- '조선일보' 등과 같이 신문사(OGG_MEDIA)와 출판물(AF_WORKS)이 혼용되는 경우에는 OGG_MEDIA 로 할당
- '스카이라이프'와 같이 방송관련 기업도 OGG_MEDIA 로 할당
- '불교방송'과 같은 종교 관련 방송기관도 OGG_MEDIA 에 해당됨
- 전문지, 잡지 등도 OGG_MEDIA 에 해당됨 (예: 애니메이션 전문지 '애니메이션툰', 영화잡지 '씨네 21')

5.5 OGG_SPORTS

- 정 의

- 스포츠 기관/단체

- 예 제

- 국민체육진흥심의위원회, 국기원 등
- 국민생활체육협의회, 국군체육부대, 한국빙상지도자협의회

- 본 문

- 1991 년 1 월 사단법인 국민생활체육협의회 창립총회를 거쳐 정식 출범하였다.
- 1984 년 체육지도대를 모태로 국군체육부대를 재창설하고 전군(全軍)의 체육종목을 통합하였다.
- . 1996 년 8 월 10 일 《빙상회보》창간호를 발간하고, 2000 년 10 월 1 일 한국빙상지도자협의회 홈페이지를 개설하였다.

- 주 의

- “~ 대표팀”의 일반 명사도 OGG_SPORTS 로 태깅한다
예) 핸드볼대표팀, 축구대표팀
다만, 국가대표팀은 개체명 대상이 아님. 앞에 특정한 스포츠 유형이 들어가야 개체명 대상이 됨 ● 축구 국가대표팀, 축구 대표팀 등
예) 우리나라 축구 국가대표팀/OG (O), 우리나라 축구 국가대표팀/OG (X)
남북/LC 공동 올림픽/EV 아이스하키 대표팀/OG (O)
- 1 부(리그) 등은 '잉글랜드 3 부 리그'로 구체화된 경우에만 OGG_SPORTS 로 할당
- 우승팀, 준우승팀의 일반명사는 태깅 대상에서 제외
- “~ 리그”의 경우는 OGG_SPORTS 로 할당
예: K 리그, 메이저리그, 프리미어리그

5.6 OGG_ART

- 정 의

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

- 예 제

- 대한극장, 구겐하임미술관, 국립중앙박물관 등

- 예술의 전당, 한국민속박물관, 로스앤젤레스필하모니관현악단, 동양극장
- 본 문
 - 예술의 전당은 LG 유플러스와 문화예술 콘텐츠 보급 독점 업무협약을 맺었다.
 - 1975 년 4 월 11 일에 마침내 한국민속박물관의 개관을 보았다
 - 1919 년 윌리엄 A.클라크 2 세(William A. Clark, Jr.)가 창단한 로스앤젤레스필하모니 관현악단은 예술제와 강연에도 참여하고 있으며, 모든 연령층의 청중에게 무료 음악회를 열어 주기도 한다.
 - 따라서 동양극장에는 많은 연극인이 모여들어 1930 년대와 1940 년대에는 대중연극의 메카로 군림했는데,
- 주 의
 - 문맥에 따라 기관이 아닌 건축물의 개념으로 사용되었다고 판단되는 경우, OG 가 아닌 AF 로 태깅한다.

5.7 OGG_MEDICINE

- 정 의
 - 의학/의료 기관/단체
 - 병원, 약국 등
- 예 제
 - 세계보건기구 등
 - 국립재활원, 경희의료원, 대한간호협회, 의료보험연합회
- 본 문
 - 1986 년 8 월 18 일 국립재활원으로 직제를 개정하였으며, 1994 년 1 월 29 일 원내에 재활병원을 설치하였다.
 - 경희의료원은 동서의학의 상호보완과 접목을 통한 '제 3 의학' 창조에 심혈을 기울이고 있다.
 - 1948 년 대한간호협회로 개칭하였고, 1949 년에는 국제간호협의회(ICN)에 정회원국으로 가입하였다.
 - 1988 년 의료보험연합회로 명칭을 변경하여 공·교 공단의 진료비 심사기구를 흡수 통합하였으며,

5.8 OGG_RELIGION

- 정 의
 - 종교 기관/단체 및 종파
- 예 제
 - 증산교, 구가톨릭교, 도교, 불교, 가톨릭 등
 - 대한성공회, 한국기독교장로회, 대한불교화엄종, 조선불교청년회

- 본 문

- 그 뒤 증산교는 보천교(普天敎)·미륵불교·모악교(母岳敎)·증산대도교·제화교(濟化敎)·태을교(太乙敎) 외에 도리원파(桃李園派) 교단, 문공신과 김병선의 교단 등으로 분열되었다.
- 복고가톨릭교회라고도 한다. 원헌에서 구가톨릭교도 대회를 열고 독립교단을 발족시켰으며, 오스트리아·스위스·프랑스·이탈리아 등 여러 나라에도 구가톨릭 교회가 생겨났다.
- 제 3 대 주교 조(趙)마가는 1911 년 교구장(敎區長)이 되어 대한성공회의 체제를 갖추고, 서울 정동에 로마네스크양식의 대성당을 건립하는 한편 선교정책을 확립하였다.
- 1961 년 18 개 노회를 9 개 노회로 통합하고, 교단명을 '한국기독교장로회'로 개칭하면서 새로운 도약기를 맞았다.
- 이 두 절의 주지를 겸임하면서 1965 년 10 월 대한불교화엄종을 창종(創宗)하고, 1966 년 6 월 불교재산관리법에 따라 불교단체 등록을 함으로써 정식으로 발족하였다.
- 1948 년 11 월 11 일 조선불교청년회가 재창립하고 1962 년 6 월 20 일 대한불교청년회로 개칭하여 오늘에 이른다.

- 주 의

- "구가톨릭교도"와 같이 종교 기관이 복합명사의 일부이고, 종교 기관만을 독립적인 의미로 분리하기 힘든 경우에는 OGG_RELIGION 으로 태깅하지 않는다.
- 종교/종파의 교단과 같은 경우에는 OGG_RELIGION 으로 태깅하지 않는다.
- 유교는 종교 단체가 아니라 철학사상(FD_PHILOSOPHY)으로 구분함

5.9 OGG_SCIENCE

- 정 의

- 과학 기관/단체
- 연구소, 연구원

- 예 제

- 국제전기통신연합, 고등과학원 등
- 농지개량조합, 한국전기안전공사, 정보통신연구진흥원

- 본 문

- 약칭은 ITU 이다. 1865 년 5 월 17 일에 국제전신연합으로 창설하여 1932 년 마드리드 만국무선전신회의에서 지금의 명칭으로 바꾸었다.
- 농지개량조합은 농지개량의 목적을 달성하기 위하여 조합구역 내의 농지개량 시설을 효과적으로 유지·관리하고
- 1990 년 4 월 특수법인 한국전기안전공사로 재발족하고, 1995 년 1 월 전기안전시험 연구원을 개원하였으며 1995 년 7 월 재난관리 책임기관으로 지정되었다.
- 1999 년 법정법인인 정보통신연구진흥원으로 새로 출범한 뒤 그해 해외정보통신기술협력센터와 부설 정보통신기술이전센터를 신설하였다.

- 주 의
 - "한국정보과학회"와 같은 과학 관련 학회도 OGG_SCIENCE 로 태깅한다.

5.10 OGG_LIBRARY

- 정 의
 - 도서관 및 도서관 관련 기관/단체
- 예 제
 - 국립중앙도서관, 바티칸도서관, 알렉산드리아도서관 등
 - 국회도서관, 한국도서관협회, 대동서관, 바티칸도서관
- 본 문
 - 국회도서관의 주요 발간자료이던 《정기간행물기사색인》《한국박사 및 석사학위논문총목록》의 발간을 중단하고,
 - 1955 년 한국도서관협회가 재발족하면서 도서관 교육문제가 거론되고, 대학에서도 도서관학을 교과과정에 넣게 되었으며, 1957 년부터 도서관학이 대학교의 독립된 과(科)로 인정받게 되었다.
 - 사립공공도서관의 효시는 1906 년에 설립된 평양의 대동서관(大同書觀:주문옥 ·곽용순 ·김흥윤이 설립)이며,
 - 1471 년부터는 바티칸도서관을 확장하는 등 학문과 예술을 장려함으로써 문화사상(文化史上) 위대한 업적을 남겼다.

5.11 OGG_LAW

- 정 의
 - 법률 기관/단체
- 예 제
 - 대한법률구조공단, 광주지방법원 등
 - 대한법무사협회, 저작권심의조정위원회, 대한변호사협회, 헌법재판소
- 본 문
 - 1896 년 대한제국 법부 전라북도재판소로 출범해 광주지방법재판소를 거쳐 1912 년 조선총독부 광주지방법원, 1945 년 미 군정청 광주지방법원, 1947 년 광주지방심리원으로 바뀌었다가 1949 년 광주지방법원으로 환원하였다.
 - 법무사법에 의하면 지방법무사회는 연합하여 회칙을 정하고 대한법무사협회를 설립해야 하며, 협회는 법인으로 하도록 되어 있다.
 - 저작권에 관한 사항을 심의하고 분쟁을 조정하기 위해 저작권 심의조정위원회를 둔다.
 - 변호사로서 개업을 하고자 하는 때에는 지방변호사회를 거쳐 대한변호사협회에 등록을 하여야 한다.

- 헌법재판소는 법적 분쟁을 사법적으로 해결한다는 점에서는 일반법원과 같으나 특히 정치적 파급효과가 큰 헌법적 분쟁을 대상으로 하며,

- 주 의

- 법원은 국가기관(OGG_POLITICS)과 법률기관(OGG_LAW)에 모두 해당되는데, OGG_LAW 로 태깅한다.
- 대한제국 법무 전라북도재판소 → <대한제국:LCP_COUNTRY> <법부:OGG_LAW> <전라북도재판소:OGG_LAW>
- 미 군정청 광주지방법원 → <미군정청:OGG_POLITICS> <광주지방법원:OGG_LAW>

5.12 OGG_POLITICS

- 정 의

- 정부/행정 기관, 공공기관, 정치기관
- 대사관

- 예 제

- 정부/행정 기관: 대사관, 미래부, 정보통신부, 우체국, 경찰서, 소방서, 동사무소, 전 화국, 시청, 군청, 국회, 대사관, 경제기획원, 북방경제정책 실무위원회, 대검, 검찰, 검찰청, 정부
- 공공기관: 공단(예: 근로복지공단, 사립학교교직원연금공단), 공사(예: 인천국제공항공사), 진흥원(예: 한국사회적기업진흥원)
- 정치기관: 새누리당, 새천년민주연합, 민주당, 독립협회
- 신궤조성도감, 기획예산처, 행정자치부, 경찰청, 스웨덴대사관

- 본 문

- 1394 년(태조 3) 9 월 신궤조성도감(新闕造成都監)을 두고 청성백(靑城伯), 심덕부(沈德符), 좌복야 김주(金湊), 전정당문학 이염(李恬), 중추원학사 이직(李穰) 등을 판사에 임명하여 실무를 담당
- 그러나 1999 년 5 월 정부조직법이 다시 개정되어 기획예산처가 신설되면서 예산청은 기획예산처의 예산실로 축소 개편되었다.
- 개정으로 내무부에서 행정자치부로 명칭이 변경되었다.
- 경찰의 정치적 중립을 지향하는 경찰청으로 발족하였다.
- 먼저 1963 년 주(駐)스웨덴대사관이 설치되었고, 1979 년 주한 스웨덴대사관이 설치되었다.

- 주 의

- 독립운동 기관도 일종의 정치기관으로 고려해서 OGG_POLITICS 로 태깅함 (예: 간도 국민회, 간도대한민회). 단, 독립운동 단체 중에 전투 단체는 OGG_MILITARY 에 해당됨
- 국가 연구기관은 OGG_POLITICS 에 해당됨 (예: 집현전). 국가 연구기관이지만 과학 기관은 OGG_SCIENCE 에 해당됨
- '경찰'은 행정기관으로 OGG_POLITICS 를 의미하기도 하고, 경찰 공무원으로 CV_OCCUPATION 을 의미하기도 한다. 사전적으로 조직을 의미하므로 '경찰'은 OGG_POLITICS 로 태깅함
- 서론, 노론 등 조선시대 정치 집단은 OGG_POLITICS 로 태깅
- 한국 정부 : 한국/LC+정부/OG, 이명박/PS+정부/OG 로 태깅한다.
- LC 와 OG 가 연이어 나올 경우, LC 가 단순히 기관이 소재한 장소를 의미할 경우에는 각각 태깅한다. 그러나 LC 를 포함하여 하나의 기관 이름으로 기능할 경우에는 전체를 함께 OG 로 태깅한다.
 - ◆ 예) '광주교육청'의 경우 '교육청'이 '광주'에 위치한다는 의미가 아니라 '광주' 지역의 교육 행정을 책임지는 기관이라는 의미로서 전체가 하나의 기관으로 인식되기 때문에 통합하여 태깅한다.
 - ◆ 예) '미국 컬럼비아 대학'의 경우 '컬럼비아 대학'이 '미국'에 위치한다는 뜻이므로 '미국/LC', '컬럼비아 대학/OG'와 같이 나누어 태깅한다.
 - ◆ '한국 정부'와 '광주교육청'의 태깅 방식이 다른 것은 띄어쓰기와 관련이 없음. '광주교육청'의 경우 개체의 실체가 명확하며, 문맥에 따라서 AF 로도 볼 수 있음. 그러나 '한국 정부'는 개체의 실체가 명확하지 않고 'OO 정부'처럼 여러 정부의 종류를 지칭하는 것으로 많이 쓰이기 때문에 '한국/LC+정부/OG'로 태깅한다.
- 양구군은 ... (중략) ... 군은 → 양구군/OG, 군/OG 과 같이 '시, 군, 구'가 줄임 표현으로 나왔을 경우 '군'도 태깅한다. 즉, '시, 군, 구'도 기관으로 나올 시, OG 로 태깅한다.
- '시, 군, 구'는 단독으로 나와도 기관의 의미를 가질 경우에는 OG 로 태깅한다.
- '대한민국 상해 임시정부'의 경우 대한민국/LC+상해/LC+임시정부/OG 로 태깅한다.
- 국가명은 맥락에 따라 LC 가 될 수도 있고 OG 가 될 수도 있는데, 그 판단 기준은 해당 문장의 해당 문장의 서술어를 중심으로 한, 문장 의미에 따라 결정한다.
 - 예) 북한은 한반도 비핵화를 위해 노력하고 있다며 → 북한/OG
- 단, '명사(국가/도시/행정구역명/)+명사' 구성에서 선행하는 '명사(국가 등)'는 LC 로 일괄 태깅한다.
 - 예) 서울 시장 → 서울/LC, 시장/CV

■ 전문분야(법률)

- 임시회, 정기회는 기관과 연관된 단어가 아니며, 특정한 의미를 가지지 않는 포괄적인 단어이기 때문에 개체명 대상이 아님.
- 합의제행정기관, 하급행정기관은 특정한 의미를 가지는 기관이 아니므로 개체명 대상이 아님

5.13 OGG_FOOD

- 정 의
 - 음식 관련 업체 및 회사
- 예 제
 - 맥도날드, 김밥천국, 맛나분식
- 본 문
 - 창밖으로 보이는 독일의 맥도날드, 우리가 1 달동안 이 맥도날드를 그렇게나 많이 가게 될 줄은 몰랐어요.
 - 우선 저희집은 경기도 포천에서 작은 "바위식당"이라는 식당을 운영하고 있습니다.
 - 타지역 수리매장의 향의 전화와 협박 죽고싶을 만큼은 고통을 견디지 못하고 문을 닫고 도미노피자에서 피자 배달을 하고 있다.

5.14 OGG_HOTEL

- 정 의
 - 숙박 관련 업체 및 회사
 - 호텔, 모텔, 민박, 펜션, 콘도, 리조트
- 예 제
 - 신라호텔, 한화콘도, 대명리조트
- 본 문
 - 단 이번 주말에는 휘닉스파크에서만 가능하고 다음주에는 수도권 지역에서만 가능합니다.
 - 좌측이 한화리조트 마린타워 지역 요즘 부동산 가격이 엄청 올라간다고합니다.
 - 흥사관광단지를 지나 유선여관과 대흥사 파안교를 건너면 오른쪽 언덕에서 석간수가 솟아나고 왼쪽으로 대흥사란 석주가 우뚝 솟아 보인다.
 - 이용대-이효정 조는 17 일 오전 11 시 30 분 중구 소공동 웨스틴조선 호텔 그랜드볼룸에서 열린 제 14 회 코카콜라 체육대상 시상식에서 우수선수상을 수상했다

5.15 OG_OTHERS

- 정 의
 - 기관/단체의 세부 분류에 해당되지 않는 나머지 모든 기관/단체
- 예 제
 - 유색인종향상전국협회, 국제소비자연맹
 - 육영재단, 녹색연합, 참여연대, 한국정신대문제대책협의회
 - 게임 관련 기관: 길드, 클랜, 파티 (게임회사는 OGG_ECONOMY 에 해당됨)
- 본 문
 - 이러한 상황에서 20 세기 초에 들어와 1909 년 유색인종향상전국협회(National Association for Advancement of Colored People:NAACP)의 결성을 비롯한 전국 도시 동맹이나 그 밖의 흑인해방운동 조직이 만들어졌다
 - 국제소비자기구라고도 한다.
 - 육영재단이란, 넓은 의미에서는 후세 교육을 위해 설립한 재단에 대한 총칭으로 쓰이기도 한다.
 - 녹색당 창당준비위원회가 배달녹색연합으로 통합되고, 1996 년 4 월 명칭이 녹색연합으로 바뀌었다.
 - 결성 당시에는 '참여민주사회와 인권을 위한 시민연대'였으나 두 차례 변경을 거쳐 참여연대로 확정되었다.
 - 1990 년 11 월 16 일, 일본군 종군위안부(정신대)에 대한 진실규명과 어렵게 사는 일본군 종군위안부 생존자들을 지원하려는 37 개 여성·시민·종교·학생단체들이 모여 만든 연합단체이다. 줄여서 정대협이라 한다.
- 주 의
 - '비세그라드 그룹' 등의 세부 분류가 애매한 협력체는 OG_OTHERS 로 태깅함
 - '유네스코'는 교육, 문화, 과학의 여러 분야와 관련된 기관이므로 OG_OTHERS 로 태깅함.

6. LOCATION

LOCATION 은 지역/장소와 지형/지리 명칭 등을 모두 포함한다. 지역/장소는 LCP(LoCation Place) 계열로 표시하고, 지형/지리 명칭은 LCG(LoCation Geography) 계열로 표시한다. 기타 지역으로 관광명소는 LC_TOUR, 천체 명칭은 LC_SPACE 로 태깅한다.

- 주 의

■ 지역/장소

- ◆ 지역명으로 주로 공식적 지역명칭과 행정구역 명칭을 나타낸다.
- ◆ 지역명은 세부적으로 분리하여 태깅한다.

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

■ 지형/지리 명칭

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

6.1 LCP_COUNTRY

- 정 의

- 국가명, 왕국

- 예 제

- 대한민국, 일본, 중국, 미국, 영국 등
- 터키, 그리스, 짐바브웨, 케냐

- 본 문

- 동부의 리제는 터키 전지역 중에서 가장 강수량이 많아 연평균 1,788mm 에 달한다.
- 예술은 그리스의 생활을 잘 반영하고 있다. 여름이면 고대 극장에서 그리스의 연극을 상연한다.
- 짐바브웨는 영국연방에 속한 국가로서, 새 헌법에 따르면 정체(政體)는 공화제이다.
- 케냐의 사회는 여러 인종이 혼재하는 다인종(多人種) 사회의 특색을 지니고 있다.

- 주 의

- 민주공화국은 특정한 나라가 앞에 붙어 있는 경우만 개체명 대상으로 인식
 - ◆ 예) 민주공화국 => 개체명 대상 아님, 콩고민주공화국 => LCP_COUNTRY

6.2 LCP_PROVINCE

- 정 의
 - 도, 주 지역명
- 예 제
 - 경기도, 강원도, 캘리포니아주, 애리조나주 등
 - 충청남도, 퀘벡주, 슈비츠주, 네바다주
- 본 문
 - 전국의 행정구역을 13 도로 구획하면서 충청남도와 충청북도로 나누어, 대체로 현재와 같은 행정구역의 윤곽을 지니게 되었다.
 - 이 퀘벡주가 캐나다 연방에서 이탈하여 독립하려는 운동이 1960년대부터 일기 시작하여 1980년에 처음으로 분리를 둔 주민의 찬반투표가 이루어졌으나 반대 60%로 무산되었다.
 - 스위스 중부에 있는 슈비츠주의 주도이며,
 - 최대산업은 관광산업으로, 네바다주는 거의 모든 관광·오락에 대해 장려하며, 각지에 인공적인 관광·오락시설이 발달되었다.
- 주 의
 - 나라에 따라서 행정구역의 세분화된 단위가 다르므로, 도시보다 큰 단위의 모든 행정구역은 LCP_PROVINCE 로 태깅한다.
 - 중국이나 일본의 지역명도 태깅 범위에 포함된다. '길림성, 산둥성'도 우리나라의 '도'에 해당하는 행정 지역이기 때문에 LC 로 태깅한다.
 - 뒤에 지방이 붙으면 지방을 제외하고 태깅을 한다.
 - ◆ <호남:LCP_PROVINCE> 지방

6.3 LCP_COUNTY

- 정 의
 - 군, 면, 읍, 리, 동 등과 같은 세부 행정구역명
- 예 제
 - 영월군, 도암면, 횡계리, 노암동 등
 - 나주군, 완도군, 담양읍, 춘포면, 영남리
- 본 문
 - 고려·조선시대를 통하여 주로 목(牧), 부(府)로서 전남지역 행정의 중심지를 이루었으며, 1895년 나주군이 되었다.
 - 1896년 영암·강진·해남·장흥의 4개 군에 속했던 도서를 통합하여 완도군이라 칭하였다.
 - 담양읍 일원에서는 가내수공업으로 죽세공예품을 생산한다.
 - 서쪽으로 공덕면(孔德面), 남쪽으로 용지면(龍池面), 북쪽으로 익산시 춘포면(春浦面)에 접한다.
 - 1952년 12월 북한의 행정구역 개편 때 면이 폐지되면서 통합되어 신설한 은천군 영남리가 되었다가 1974년 5월 남포직할시에 이관되었다

- 주 의

- 2014 년부터 정식 시행된 새로운 주소 표기법인 도로명주소법에는 도로명이 주소로 표시된다. 도로명주소법에서의 도로명은 AF_ROAD 로 태깅함
- 나라에 따라서 행정구역의 세분화된 단위가 다르므로, 도시보다 작은 단위의 모든 행정구역은 LCP_COUNTY 로 태깅한다

6.4 LCP_CITY

- 정 의

- 도시명

- 예 제

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

- 본 문

- 현재 맨체스터는 공업도시보다는 상업·금융·보험 등의 상업도시로서 런던에 버금가는 도시가 되었다.
- 제임스 W.덴버(James W. Denver)의 이름을 따 덴버로 지명을 바꾸었다.
- 혼모쿠곶과 오오카강 하구 사이의 천해(淺海)가 매축되어 현재의 요코하마의 중심부가 형성되고,
- 부에노스아이레스는 1536 년에 창건되었으나 인디오의 공격을 받아 1541 년에 방치되었다가, 1580 년에 재건되었다.
- 터키 최대의 공업도시이기도 한 이스탄불은 섬유·식품가공·농기구·고무 등의 공업이 발달하였으며 군수공업도 중요한 몫을 차지한다.

- 주 의

- 국가의 수도명은 LCP_CAPITALCITY 으로 태깅한다.

6.5 LCP_CAPITALCITY

- 정 의

- 수도명

- 예 제

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

- 본 문

- 아테네는 고대 그리스 아티카 지방의 중심지였다.
- 1790 년 국회의 동의를 얻어 미국의 수도로 지정되었고, 1878 년 조지타운이 워싱턴의 일부가 된 이래 새로운 국가의 영구 행정부 소재지로 선정된 컬럼비아 특별구와 공존하고 있다.
- 수도 캔버라와 주변지역으로 이루어져 있다.
- '수도에의 집중현상'은 프랑스의 특징이며, 파리는 세계 제 4 위의 인구밀집 지역이다.

- 브라질리아를 21 세기의 도시로 건설하였다.

- 주 의

- 현재는 일반 도시이지만 예전에 수도였던 도시는 LCP_CITY 로 태깅한다.
◆ 예) 아르마비르, 음반자콩구
- 현재는 없어진 도시이지만 예전에 수도였던 도시는 LCP_CITY 로 태깅한다.

6.6 LCG_RIVER

- 정 의

- 강/호수/육수 이름

- 예 제

- 섬진강, 대동강, 한강 등
- 충주호, 동강, 춘천호, 아산호

- 본 문

- 기암괴봉이 거대한 병풍처럼 펼쳐지면서 충주호와 어우러져 뛰어난 경관을 연출한다.
- 동강은 완택산(完澤山:916 m)과 곰봉(1,015 m) 사이의 산간지대를 감입곡류하며 남서쪽으로 흐르다가 영월군 영월읍 하송리(下松里)에서 서강(西江) 만나 남한강 상류로 흘러든다.
- 춘천호에 담긴 물은 북한강 상류 지역을 호수화하여 멀리 화천군 파로호(破虜湖) 부근에까지 미치고 있는데,
- 아산호는 남양호와 함께 1977 년에 국민관광지로 지정되어 유원지로 개발되었으며, 수로(水路)는 낚시터이다.

6.7 LCG_OCEAN

- 정 의

- 해양/바다 명칭

- 예 제

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

- 본 문

- 한국에서는 조선시대에 조선해(朝鮮海) 또는 창해(蒼海)라고 하였으나, 8·15 광복 후부터 동해라 부르기 시작하였다.
- 이어 1521 년 세계일주에 나선 마젤란이 대서양에서 남아메리카 남단의 마젤란 해협을 지나 태평양을 거쳐 필리핀에 이르렀다.
- 인도양에서는 거의 그 중앙부에 인도양 중앙해령이 남북으로 뻗어 있고,
- 산호해(珊瑚海)의 일부를 형성하며 서쪽 끝은 토러스 해협을 거쳐 아라푸라해(海)로 통한다.

- 주 의
 - 바다의 일부분을 뜻하는 해역도 LCG_OCEAN 에 해당됨 (예: 바람내리포)

6.8 LCG_BAY

- 정 의
 - 반도/만의 명칭
- 예 제
 - 도쿄만, 라몬만, 랴오둥만, 랴오둥반도 등
 - 아산만, 울산만, 변산반도, 발칸반도, 이베리아반도
- 본 문
 - 충청남도 아산시와 경기도 평택시 사이에 있는 아산만은 밀물과 썰물의 차이가 심한 곳인데
 - 울산만 지역은 옛날부터 어업이 성하여,
 - 황해연안에 돌출한 변산반도는 지괴산지의 대표적인 것으로
 - 때문에 대체로 발칸 반도의 기후는 건조하며, 식생(植生)은 특히 남부에서 빈약하다.
 - 포르투갈은 1580 년부터 에스파냐의 속령이 되었으며, 1640 년에 독립을 회복한 후로는 이들 양국이 이베리아 반도를 차지하고 있다
- 주 의
 - '간절곶', '호미곶', '데즈네프 곶' 등의 곶도 LCG_BAY 에 해당됨

6.9 LCG_MOUNTAIN

- 정 의
 - 산맥/산, 광산, 봉우리의 명칭
- 예 제
 - 라우산, 라웅산, 라키화산, 래밍턴산 등
 - 태백산맥, 차령산맥, 계룡산, 금강산, 대청봉
 - 광산 이름 : 상동광산
 - 봉우리 이름 : 인수봉, 용암봉, 노적봉 등
- 본 문
 - 태백산맥은 기후풍토상으로도 영동(嶺東)·영서(嶺西) 양지방의 경계선을 이루며,
 - 또 차령산맥은 최한월(最寒月) 평균기온 - 3 ℃의 등온선과 일치하여 한국의 기후 구를 남부의 온대와 북부의 냉대로 크게 구분하는 경계가 되고 있다.
 - 전체 능선의 모양이 마치 닭뿔을 쓴 용의 형상을 닮았다 하여 계룡산이라고 불린다.
 - 해금강이란 이름은 경치가 금강산을 닮았다고 하여 남택하(南宅夏)가 지은 것이라 고 한다.
 - 최고봉은 대청봉이다. 대청봉 남쪽에 한계령, 북쪽에 마등령·미시령 등의 고개가 있다.

6.10 LCG_ISLAND

- 정 의
 - 섬/제도 명칭
- 예 제
 - 나이팅게일섬, 노시베섬, 로드리게스섬, 로스제도, 마스카렌제도 등
 - 대난지도, 대부도, 백령도
- 본 문
 - 대난지도의 옆의 작은 섬이라 하여 소난지도라는 이름이 붙었다
 - 서해안에서 제일 큰 섬으로, 큰 언덕처럼 보인다고 하여 대부도라고 하였으며,
 - 따오기가 흰 날개를 펼치고 공중을 날르는 모습처럼 생겼다 하여 백령도라고 한다.
- 주 의
 - '제주도'는 섬의 이름이기도 하지만, 행정구역으로 '제주특별자치도'를 의미하기도 하므로 LCP_PROVINCE 로 태깅한다.
 - '섬'이면서 '주'인 경우, LCP_PROVINCE 로 태깅 (예: 코스라에주)
 - 환초는 산호섬이므로 LCG_ISLAND 로 태깅 (예: 윗제환초)

6.11 LCG_CONTINENT

- 정 의
 - 대륙 명칭
- 예 제
 - 아메리카, 아시아, 오세아니아, 유라시아 등
 - 유럽, 북아메리카, 남아메리카, 아프리카
- 본 문
 - 유럽은 육반구(陸半球)의 중앙에 위치하고 있어 세계의 다른 대륙에 쉽게 도달할 수 있는 편리한 위치에 놓여 있다.
 - 북아메리카 대륙은 대체로 멕시코 남부를 가로지르는 테우안테펙 지협(地峽)을 남쪽 한계로 하고 있으며,
 - 대륙으로서 아프리카라는 이름으로 불린 것은 16·17 세기 네덜란드의 항해자들이 이곳이 독립된 대륙이라는 사실을 알고 난 뒤부터이다.

6.12 LC_TOUR

- 정 의
 - 관광명소
 - 온천, 해변, 관광지+해안, 식물원, 수목원, 정원, 공원, 세트장, 광장, 테마파크, 농원/농장, 폭포, 계곡, 동굴, 야영장, 체험장, 수련장, 샘터, 약수터
- 예 제
 - 간현유원지, 대성산유원지, 디즈니월드, 디즈니랜드, 서울랜드
 - 다도해 해상국립공원, 에버랜드, 예송리해수욕장, 계룡산국립공원
 - 해운대해안, 양남해변, 백암온천, 여미지식물원, 국립수목원, 미소정원, 울산바위, 노안청소년야영장, 이안숲속, 용정약수터
- 본 문
 - 다도해 해상국립공원은 다도해 여러 곳에 지역별로 나뉘어 지정되어 있다.
 - 에버랜드의 새로운 놀이공원으로 물을 주제로 한 테마파크이다.
 - 대체로 키작은 나무로 형성된 계룡산국립공원의 삼림지대는 자연경관이 뛰어나 수려한 산봉이 15 개, 계곡이 7 개소, 폭포가 3 개, 이름난 동굴 5 개소가 있다.
- 주 의
 - ‘올드페이스풀 간혈천’ 등과 같은 간혈천도 온천이므로 LC_TOUR 에 해당됨

6.13 LC_SPACE

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

- 본 문

- 태양은 모든 천체 가운데 가장 두드러진 존재이고,
- 시리우스의 동반성은 처음으로 발견된 백색왜성이다.
- 그 대표적인 예로 오리온자리 말머리성운·백조자리 북아메리카성운·남십자자리 석탄부대(coalsack) 등이 있으나, 은하면에 따른 영역에 많다.
- 장미성운은 내부의 Oe 형에 의하여 자극되어 휘선을 발하는 발광산광성운이다.
- 미자르는 1857 년 G.P.본드가 처음으로 쌍성의 사진촬영에 성공한 별로도 알려져 있다.
- 토성의 표면에는 띠 모양의 무늬가 보이는데,
- 지금까지 화성과 목성 사이에서 7 천여 개의 행성이 발견되었다. '세종'별도 그 중의 하나이며, 지름 5~6km 의 작은 행성이다.
- 지구에서 가장 멀리 떨어져 있는 명왕성(冥王星)까지의 거리는 약 60 억km, 즉 빛이 5.5 시간 동안에 가는 거리이다.
- 미란다를 비롯한 천왕성의 위성들도 천왕성의 자전면과 거의 일치하는 궤도를 가지고 있다.
- 1759 년 헬리혜성의 회귀(回歸)를 프랑스에서는 처음 관측한 후, 혜성연구에 전념하여 15 개의 새로운 혜성을 포함하여 21 개의 혜성을 검출하였다.
- 사자자리의 나머지 별은 데네볼라의 서쪽에 있는데, 6 개의 별이 마치 물음표(?)를 뒤집어놓은 형태로 되어 있어 쉽게 찾을 수 있다.
- 1872 년 아르헨티나에서 남쪽하늘을 관측하던 B.A.골드가 나침반자리로 부를 것을 제안하여 1922 년 국제천문연합에서 공인되었다.
- 별자리의 물병자리는 가니메데스가 물병을 들고 있는 모습이라고 하며, 그 근처에 제우스가 독수리가 되어 날아가는 모습이라고 하는 독수리자리가 있다.
- 오리온자리를 바라볼 때 오른쪽의 어깨 부분에 있는 청백색의 밝은 별이다.

6.14 LC_OTHERS

- 정 의

- 기타 지역/장소
- LC 계열의 세부유형이 아닌 기타 장소
- 지하철/기차역, 공항, 버스터미널, IC, 시장, 발전소, 정류장, 단지, 항구, 골짜기

- 예 제

- 방화역, 김포공항, 인천국제공항, 남부터미널, 태안화력발전소, 모란시장, 부산항
- 북해유전, 북자파라나 유전
- 서양, 동양, 동구, 서구, 북방, 서방, 남방, 동방
- 가로구릉, 기아나고지, 나가구릉 등
- 미시시피삼각주, 동아프리카대지구대
- 남대문 정거장, 서울역 광장
- 적도
- 언덕, 나루, 금광, 사구, 산업단지 이름 (구체적인 명칭이 나왔을 때 한정하여 태깅함)

- 본 문

- 일몰과 일출을 한 자리에서 볼 수 있어 많은 관광객들이 찾고 있는 충남 서천 마량 포구에서 올해를 마무리하는 풍성한 이벤트가 펼쳐진다.
- 축제의 주 행사인 산행은 오전 9 시 명석면 홍지마을 주차장에서 모여 준비운동으로 몸을 풀고 출발하며 해발 420 의 광제봉수대에 올라 주변 경관을 감상한 뒤 홍지마을까지 되돌아 오는 13 구간에서 실시된다.
- 그러다가 5000 년 전부터 삼각주의 퇴적환경이 만들어지고, 현재의 미시시피 삼각주가 이루어졌다.
- 오래전부터 잘 알려진 것은 동아프리카대지구대로, 남단은 모잠비크에서부터 북단은 요르단 근처까지 길이 6,000km 에 이르며, 그 내부에서 여러 분지(分枝)로 갈라져 있다.

- 주 의

- ~사막, ~오아시스는 LC_OTHERS 에 해당됨
- 해안, 해변(ex. 서해안, 해운대해안, 양남해변): 해변은 LC_TOUR, 해운대와 같이 관광지 해안은 LC_TOUR, 서해안/남해안 등은 LC_OTHERS
- 제주항/목포항: 만/해안이 아닌 항의 경우는 LC_OTHERS
- 슬로프명(이스턴 슬로프, 루키 슬로프): LC_OTHERS
- 시장명 (ex. 자갈치시장, 가락시장, ...): LC_OTHERS
- ~검문소(배둔검문소)/무기제작소/군기창고/군사보호구역/군사훈련장: 개체명인 경우에는 LC_OTHERS 에 해당됨
- 구체적인 유적지, 마을, 터, 성지, 전승지 등은 문화재이면 AF_CULTURAL_ASSET 이고, 문화재가 아니면 LC_OTHERS 에 해당됨
- '강', '바다', '피오르', '대륙붕', '간석지', '우발레', '피요르드', '주상해분' 등의 단순한 지형 명칭은 개체명 대상이 아님. 나일강, 송네 피오르드 등의 특정 개체만을 LC 계열의 개체명으로 태깅함
- '게토' 등의 특정 구역을 가리키는 것은 LC_OTHERS 에 해당됨
- '언덕'은 관광지이면 LC_TOUR 이고, 관광지가 아니면 LC_OTHERS 에 해당됨 (예: 바티칸언덕)
- '나루'는 LC_OTHERS 로 해당됨 (예: 구드래나루)
- 호텔 단지는 LC_OTHERS 에 해당됨 (예: 보르시벨트)

7. CIVILIZATION

CIVILIZATION 은 문명/문화와 관련된 용어들로 구성되어 있다. 세부 문명/문화로 구분되는 것들만 대상으로 하고, 그 이외의 단어들은 대상으로 하지 않는다.

7.1 CV_NAME

- 정 의
 - 문명/문화 명칭
- 예 제
 - 인더스문명, 거석문화, 날문화
 - 황허문명, 아스테크문명, 안데스문명, 메소포타미아문명, 도식문화
- 본 문
 - 인더스문명의 대표적 유적은 당시의 2 대 도시였던 하라파와 모헨조다로인데
 - 인더스문명을 남긴 사람들의 인종적 구성도 복잡하여, 고대의 모헨조다로에서 발굴된 14 개의 인골(人骨)의 예형(例型) 중 3 예형은 원(原)오스트랄로이드, 6 예형은 지중해인종, 1 예형은 몽골인종, 4 예형은 알프스 인종이었다
 - 입석은 선사시대 거석문화의 한 형태로 지역 간의 경계를 나타내거나 액운을 막아 주는 신앙의 대상으로 세워졌으며
 - 암리(Amri) ·날문화라 병칭되는 바와 같이 암리문화와 많은 연관이 있는데, 분포범위와 가지무늬토기[彩文土器]의 양식이 약간 다를 뿐이다
 - 은나라의 청동기는 매우 특이한 것으로서 황허문명의 상징이다.
 - 이처럼 발달한 아스테크문명은 1520 년 갑자기 나타난 에스파냐의 페르난도 코르테스가 거느린 수백 명의 군대의 침입으로 수개월에 걸친 외교교섭과 격렬한 전투 끝에 멸망하였다.
 - 안데스 문명의 오랜 역사를 통하여 문자(文字)는 끝내 발명되지 않았다.
 - 메소포타미아 문명은 폐쇄적인 이집트 문명과는 매우 대조적이었다.
 - 이글루릭 에스키모 지역에서 도식문화에의 이행과정에 대한 연구가 진행되고 있다.

7.2 CV_TRIBE

- 정 의
 - 민족/종족 명칭 또한 국가를 구성하는 국민을 지칭
- 예 제
 - 이라크족, 유대인 등
 - 조선족, 마오리족, 아메리카인디언, 아리아인, 고산족
 - 집시, 보헤미안, 유목민
 - 유대인계, 히스패닉계, 한국계, 흑인, 백인

- 본 문

- 이라크족의 쿰바(Kumba)제도처럼 남녀 모두에게 배우자 이외의 사람과의 성관계를 허용하는 '성관계 범위'가 설정된 사회도 있다.
- 바빌론유수기를 거쳐 고향으로 돌아온 이스라엘인을 모두 유대인이라 불렀고
- 현재 독립운동가의 유족들을 포함한 약 200 만 정도로 추산되는 조선족은 국적이 중국인이므로 모국에의 귀환이 성사되기 어려운 현상상황에서는
- 마오리족은 아마도 동부 폴리네시아의 소시에테제도에서 쿡제도를 경유하여 몇 차례에 걸쳐 건너왔으며,
- 뒤에 본래의 인도인과 구별하기 위하여 '아메리카의 인도인', 즉 아메리카인디언이 라 부르게 되었다.
- 인도에 침입한 아리아인은 편자브 지방에 들어가 원주민을 정복하고 혼혈을 장려하였다.
- 고산족의 원주지(原住地)나 타이완으로 이주해 온 연대가 종족에 따라 상이하므로 문화수준도 종족 간에 격차가 심하다.

- 주 의

- '호모사피엔스', '호모 에렉투스', '네안데르탈인' 등도 CV_TRIBE 에 해당됨
- '~계'도 CV_TRIBE 에 해당된다. '한국계 미국인'의 경우 한국계/CV+미국인/CV 로 태깅한다.
- 흑인, 백인, 아프리카인 등 인종도 CV_TRIBE 에 해당됨

7.3 CV_SPORTS

- 정 의

- 스포츠/레포츠/레저 명칭
- 스포츠/레포츠 종목

- 예 제

- 스케이팅, 스키 등
- 하이킹, 행글라이딩, 인라인스케이팅, 축구, 마라톤

- 본 문

- 일상생활에서 교외로 벗어나는 산책 나들이에서부터 가벼운 등산 등 야외활동을 널리 하이킹이라 한다.
- 행글라이딩은 공기 속을 움직이는 스포츠이므로 기상에 민감한 편인데 기상조건으로는 바람이 특히 중요하다.
- 미국의 롤러스케이팅 스피드스케이팅 경기가 쿼드(quad)와 인라인(inline)으로 나뉘는 등 1990 년대 들어 세계적으로 인라인스케이팅 붐이 일어났다.
- 축구는 상대팀과 대결하여 득점을 겨루는 것이 목표이기 때문에 그 목표를 달성하기 위해서는 기본기술이 절대적으로 중요하다.
- 마라톤 기록은 코스마다 조건이 약간 다르기 때문에 세계기록은 공인되지 않으며, '세계 최고'라는 말이 쓰인다.

- 주의
 - "널뛰기", "칠교놀이", "놋다리밟기" 등의 놀이는 AFW_PERFORMANCE 가 아니라 CV_SPORTS 대상임
 - "널뛰기", "칠교놀이" 등이 CV_SPORTS 의 대상인 것을 참고하면 '전통놀이', '민속놀이'는 CV 의 대상에 포함되는 것으로 볼 수 있다. 따라서 '강강수월래, 사물놀이, 윷놀이, 자치기, 제기차기'는 태깅한다.
 - 하지만 달리기, 줄넘기, 공놀이, 공기놀이는 민속놀이라고 보기 어렵기 때문에 태깅하지 않는다.
 - 바둑, 당구, 포켓볼은 일상에서 게임의 형태로 많이 행해지긴 하나, 전문 종목이며 정식 스포츠에 해당되기 때문에 CV 로 태깅한다.
 - '빙고 게임'은 전문 종목, 스포츠라고 볼 수 없으므로 태깅하지 않는다.

7.4 CV_SPORTS_INST

- 정 의
 - 스포츠/레포츠/레저의 용품/도구 명칭
- 예 제
 - 스케이트, 카빙스키, 행글라이더 등
- 본 문
 - 레저용으로 흔히 볼 수 있는 롤러블레이드(roller blade)의 공식명칭인 인라인스케이트(inline skate)는 한 줄로 된 프레임에 3~5 개의 바퀴가 일렬로 고정되어 있다
- 주 의
 - 수영복, 등산복 등의 운동 의복은 CV_SPORTS_INST 가 아니라 CV_CLOTHING 로 태깅한다.

7.5 CV_POLICY

- 정 의
 - 제도/정책 명칭
- 예 제
 - 카스트, 쓰레기종량제, 직업공무원제도, 연금제도, 고시제도
 - 가부장제, 진대법, 노예제, 교과담임제, 최저임금제도
 - 교과담임제, 최저임금제도, 섯다운제도
- 본 문
 - 가부장제의 가족형태에서는 가족성원이 세습적 규칙에 따라 지명된 개인의 지배를 받는데, 대개는 장남이 세습적으로 가장의 지위와 재산을 계승하여 안으로는 가족을 통솔하고 밖으로는 가족을 대표한다.
 - 10 월에 환납(還納)하는 것을 상식(常式)으로 시행하도록 하였는데, 이것이 한국에서의 진대법 실시의 최초의 기록이다.
 - 노예제 안에는 각양각색의 사회적 신분을 가진 자가 포함된다.
 - 현재 초등학교는 학급담임제를, 중학교 이상은 교과담임제를 채택하고 있다.

- 최저임금제도를 통한 국가의 강제에 의한 임금액의 보호는 노사간에 실질적인 교섭 평등관계가 유지되고 있지 않은 사업장의 근로자들을 위해서는 절실히 요구된다.

- 주 의

- 제가회의, 화백회의, 정사암회의 등은 개체명으로 보기 어려움
- 국가 사업은 TM(TMI_PROJECT)으로 태깅한다. 정책, 제도에 한해서 CV(CV_POLICY)로 태깅한다. 따라서 '화폐정리사업'은 CV가 아닌 TM으로 태깅한다.
- 아동수당, 주택보조금, 무상급식 각각 특정한 제도, 정책이기 때문에 'CV'로 태깅한다. 그러나 교육복지 정책, 결혼 제도는 세부 제도, 정책으로 볼 수 없기 때문에 태깅하지 않는다.

7.6 CV_TAX

- 정 의

- 조세 명칭

- 예 제

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

- 본 문

- 부가가치세는 모든 재화 또는 용역의 소비행위에 대하여 부과되는 일반소비세이며
- 일용근로자의 근로소득에 대한 근로소득세는 원천징수에 의하여 과세가 종결된다
- 재산세는 건축물·선박 및 항공기를 과세물건(課稅物件)으로 하며(동법 제 180 조, 제 181 조),
- 취득세는 토지·건축물·차량·기계장비·입목(立木)·항공기·선박·광업권·어업권·골프회원권·콘도미니엄 회원권·종합체육시설 이용 회원권 등의 자산의 취득에 대하여 그 취득자에게 부과한다(지방세법 제 105 조).
- 공공단체(公共團體)가 증징(增徵)하는 세액에 대하여 부가세를 부과하도록 하였었다.

7.7 CV_FUNDS

- 정 의

- 연금, 기금, 자금, 펀드 명칭

- 예 제

- 거치연금, 문예진흥기금, 공적자금, 퇴직연금
- 노령연금, 개인연금, 통화안정기금, 방송발전기금, 공적자금

- 본 문

- 가입기간이 20 년 이상인 가입자 또는 가입자였던 자가 60 세(대통령령이 정하는 특수직종근로자는 55 세)에 달한 때에는 그때부터 그가 생존하는 동안 노령연금을 지급받는다
- 개인연금은 정부의 정책지원에 의해 저축기간 동안 연간 최고 72 만 원을 한도로 불입액의 40%가 소득공제되며, 이자소득세도 비과세되는 세제혜택을 받는다.
- 통화안정기금은 통상 중앙은행의 통화안정계정에 계상·운영된다.

- 방송법 제 36 조 방송발전기금의 설치에 관한 규정에 따라, 방송위원회가 방송진흥 사업 및 문화·예술 진흥사업을 위하여 설치 운용하는 자금을 말한다.
- 또 해외차관, 정부보유 공기업주식, 공공자금관리기금, 한국은행 등에 의해 우회적으로 투입된 자금도 넓은 의미의 공적자금으로 본다.

7.8 CV_LANGUAGE

- 정 의
 - 언어 명칭
- 예 제
 - 힌디어, 프랑스어, 이탈리아어, 산스크리트
 - 라틴어, 몽골어, 아카드어, 말레이어, 스와힐리어
- 본 문
 - 이와 같이 라틴어는 학문과 종교의 언어로서 근대 서유럽의 여러 언어에 깊숙이 침투하여 그들 언어에 많은 어휘를 제공하였다.
 - 현대 몽골어는 동서의 대방언군(大方言群)으로 나뉘고 그 밖에 약간 고립된 방언이 있으나, 지역적 언어차는 별로 심하지 않다.
 - 수메르의 설형문자에서 발전하여 설형문자 형태로 표기하는 아카드어는 약 600 개 낱말과 음절 기호가 있다.
 - 말레이어는 공통어로서 발달했기 때문에 문법에서 관용도의 폭이 상당히 넓고 또 실제 발음이나 어휘에도 지방의 차이가 심하다.
 - 스와힐리어의 모든 음절은 개음절(開音節)이다. 어순은 영어의 어순과 비슷하며 동사의 인칭변화(人稱變化)도 있다.
- 주 의
 - '외국어' 등의 일반명사는 CV_LANGUAGE 에 해당하지 않음

7.9 CV_BUILDING_TYPE

- 정 의
 - 건축양식 명칭
- 예 제
 - 기능주의건축, 노르만양식
 - 시토파건축, 에마뉘엘 양식, 그리스건축, 도리스양식
 - 유럽식 목조주택, 기와집, 널벽집, 한옥
- 본 문
 - 현존하는 가장 오래되고 전형적인 시토파건축이다.
 - 이 바타라가 에마뉘엘 양식의 중심지였다.
 - 그리스건축의 중심은 신전이므로 그 역사도 신전 형식의 발전에 비례한다.
 - 도리스 양식의 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.

7.10 CV_FOOD

- 정 의
 - 음식/국물 명칭, 음식재료
 - 음식의 유형도 포함
- 예 제
 - 쇠고기, 소금, 어리굴젓, 명란젓 등
 - 국수, 현미, 나물, 해물겨자채, 김치
 - 음식 유형: 국, 탕국, 밥, 무침, 향신료, 안주
- 본 문
 - 한국에서도 아주 오래 전부터 국수를 만들어 먹었으나, 밀의 생산이 많지 않았기 때문에 상용음식이 되지는 못하였다.
 - 현미는 정백으로 인한 영양분의 손실이 없으므로 백미에 비하여 지방, 단백질, 비타민 B₁·B₂ 가 풍부하다.
 - 나물은 그 중 속채류의 대표적인 요리라 할 수 있다.
 - 해물겨자채는 재료도 좋아야 하지만 겨자즙의 간을 잘 맞추고 넣는 양이 적당해야 맛있다.
 - 특히 계절에 따라 여러 가지 김치를 담가 먹었는데, 봄철에는 나박김치·봄배추김치·짬지를 주로 담가 먹었고, 여름에는 오이소박이·열무김치가 주류를 이루었다.
- 주 의
 - 음식 재료명으로 쓰였다고 판단되는 경우 AM 또는 PT 에 속하는 개체라 할지라도 CV 로 태깅한다. (무조건적으로 AM, PT 를 우선시하여 태깅하지 않음)
 - 생선이나 과일을 건조시킨 건어물, 건과일 등은 식용을 목적으로 식재료를 가공한 것이므로 AM_FISH 나 PT_FRUIT 가 아니라 CV_FOOD 로 태깅함 (예: 북어, 건태, 건명태, 건포도, 꽃감)

7.11 CV_DRINK

- 정 의
 - 음료수 명칭
- 예 제
 - 위스키, 우유, 요구르트
 - 주스, 홍차, 레모네이드, 칵테일, 샴페인
- 본 문
 - 주스의 원료는 오렌지·레몬·귤·사과·포도·파인애플·토마토·당근 등이다.
 - 동양 사람들은 차의 빛깔이 붉다고 하여 홍차라고 하지만 영어로는 찻잎이 검다고 하여 블랙 티(黑茶)라고 한다.
 - 레모네이드는 레몬과 에이드를 합친 말이며, 에이드란 천연과즙에 설탕과 물을 섞었다는 뜻이다.
 - 이와 같이 칵테일은 미국에서 처음 만들기 시작했다고 하나 혼성음료를 만드는 습관은 반드시 미국에서 시작된 것은 아니고
 - 등이 있다.

- 삼페인은 마개가 빠질 때 나는 평하는 소리와 함께 이는 거품이 특징인 술로서 축 하연을 할 때 쓰는 축하주이다.

- 주 의

- 태깅말뭉치에 태깅된 애매한 목록 (CV_DRINK 로 태깅함)
◆ 음료, 술

7.12 CV_CLOTHING

- 정 의

- 의복/섬유 명칭
- '티셔츠, 점퍼' 같은 의복 유형도 포함하나 '바지, 모자, 상의, 하의' 등 대분류는 태깅하지 않음.

- 예 제

- 모시, 메리야스, 와이셔츠
- 레이온, 양모, 점퍼, 카르고팬츠

- 본 문

- 레이온의 큰 결점은 내수성(耐水性)이 크게 떨어진다는 것과, 탄성이 적어 의복재료로는 구김살이 생기기 쉽다는 점이다.
- 양모는 오스트레일리아·뉴질랜드·아르헨티나·미국·러시아 등지에서 많이 생산된다.
- 오늘날과 같은 형태의 점퍼가 되었으며, 제 1 차 세계대전 후의 20 대의 여성복으로 전성기를 이루었다.
- 특히 국방색 카르고 팬츠는 카무플라주 점퍼나 셔츠를 입어 밀리터리 스타일로 연출할 수도 있다.

- 주 의

- 수영복, 등산복, 등산화 등의 스포츠 용품은 CV_SPORTS_INST 이 아니라 CV_CLOTHING 으로 태깅한다.
- 베레모, 비니, 페즈 등의 모자 종류도 CV_CLOTHING 으로 태깅한다. 그러나 '모자'만 나왔을 때는 태깅하지 않는다. ('바지'와 '모자'는 특정 의복 유형보다 더 상위층에 해당)

7.13 CV_POSITION

- 정 의

- 직위/직책 명칭
- 스포츠 포지션
- 고대 국가의 지위
- 종교인 명칭
- 신분 제도의 지위

- 예 제

- 황제, 학사, 장군, 왕, 영의정, 1 품, 대료, 귀족, 군주
- 사장, 감독자, 교주
- 4 번 타자, 유격수, 투수, 타자

- 무슬림, 기독교인, 불교인
- 노예, 백성, 주인공, 대통령
- **CV_POSITION 으로 태깅함 (애매하지만 태깅말뭉치에 태깅된 목록)**
 - ◆ 신, 여신, 악신, 선신, 연구진, 사용자, 충신, 현지인, 일반인, 사상자, 응시자, 창시자, 유권자, 국민, 주민, 저자, 실업자, 투자가, 기관투자가
 - ◆ 피서객, 포로, 피서객, 시민권자, 유부녀, 재초부, 미혼모, 강습생, 국자학생, 수험생, 연수생, 초심자, 재일동포, 동포, 대리인, 기업인, 근로자
 - ◆ 멤버, 수상자
- **CV_POSITION 으로 태깅하지 않는 목록**
 - ◆ 부호, 빈민, 처녀, 아낙네, 여자, 남자, 멘토, 관찰자

- 본 문

- 사장·부사장 등은 최고관리자, 부장·과장 등은 중간관리자, 계장·감독자 등은 하급관리자이다.
- 15 세에 음보(蔭補)로 교원(校尉)가 되고, 뒤에 장군에 이르렀다.
- 고대의 왕은 입법 ·사법 ·행정 등의 모든 국가권력을 장악하고 군사상 통수권을 가지고 있었으며 종교적으로 신격화되었다.
- 사교집단의 창설자이자 교주인 짐 존스(Jim Jones)를 비롯해 총 914 명에 달하는 신도들이 집단 자살한 사건을 말한다

- 주 의

- 프리마돈나, 소프라노, 테너 등은 CV_POSITION 으로 태깅함
- "헤이그특사"는 CV_POSITION 으로 태깅함
- '국회의원'은 CV_POSITION 과 CV_OCCUPATION 이 모두 가능하지만, CV_POSITON 으로 태깅함
- 사람들이 가질 수 있는 특정한 역할의 경우 CV_POSITION 으로 볼 수 있다. 예를 들어, 일반 사람이 텔레비전 프로그램을 보는 역할을 수행함으로써 시청자라는 개체명이 될 수 있음
- 지역명+직위, 사람이름+직위의 경우 각각 분리하여 태깅한다.
 - ◆ 중국 대사: 중국/LC+대사/CV , 류샤오밍 대사 : 류샤오밍/PS+대사/CV
 - ◆ 광주 시장: 광주/LC+시장/CV, 박원순/PS+시장/CV
- '현 사장', '전 문화예술위원회 위원장' 과 같이 직위에 '전, 현'이 포함되어 나올 경우 '전, 현'은 태깅하지 않는다. 예) 현 사장 : 사장/CV

7.14 CV_RELATION

- 정 의
 - 인간/가족 관계 명칭
- 예 제
 - 형제, 제자, 남편, 시어머니, 시누이
 - 어머니, 아버지, 친구, 딸
 - 맞벌이부부, 딩크족
 - 스승, 제자
 - 유족
- 본 문
 - 은경은 어머니의 간청으로 실어증에 걸린 아버지를 만나고 아버지에게 연민의 정을 느낀다.
 - 부모들이 친구간이거나, 친척 어린이 등과 같이 선택하려는 노력없이 일상생활에서 교섭이 빈번한 어린이들끼리 그대로 동무관계가 된다.
 - 즉 딸만 있는 집에서 데릴사위로 삼아 같이 살면서 그의 노동력(생산력)으로 생활 하는 술서(率)

7.15 CV_OCCUPATION

- 정 의
 - 직업 명칭
- 예 제
 - 화가, 극작가, 건축가, 가정교사, 기생, 언어학자
 - 공인중개사, 영양사, 군의관, 아나운서, 종군작가
- 본 문
 - 이 공인중개사는 정부가 부동산중개업자의 공신력(公信用)을 높이고 공정한 부동산 거래질서를 확립한다는 입법취지로,
 - 단체급식의 관리책임자는 영양사가 되어야 하며,
 - 부대에 배치된 군의관은 참모요원으로서 부대장병의 건강관리와 위생에 관한 직무를 수행하며 부대장을 보좌한다.
 - 아나운서는 방송국명의 고지, 방송순서의 소개, 뉴스 방송, 스포츠나 식전의 실황중계, 대담의 사회, 낭독 등 다방면에 걸쳐 방송순서를 구성하는 중요한 요소가 된다.
 - 일찍이 제 1 차 세계대전 당시 이탈리아 전선에 군속으로 종군하여 부상을 바 있는 E.M.헤밍웨이는 종군작가의 대표적인 예이다.
- 주 의
 - **애매하지만 CV_OCCUPATION 으로 태깅함 (태깅말뭉치 기준)**
 - ◆ 성직자, 농민, 선수

7.16 CV_CURRENCY

- 정 의
 - 통화 명칭
- 예 제
 - 달러, 엔, 유로, 루피, 위안, 원
- 본 문
 - 유로는 7 종의 지폐와 8 종의 동전으로 구성되며 유로의 제작·발행은 각 나라가 독자적으로 실시한다.
 - 루피는 '은'을 나타내는 산스크리트 '루피아'에서 유래한다.
 - 위안의 대외적 환율은 1950 년 7 월 중국인민은행이 환율을 공표함으로써 전국적으로 통일된 환율이 실현되었고

7.17 CV_PRIZE

- 정 의
 - 상과 훈장
- 예 제
 - 노벨경제학상, 노벨문학상, 노벨물리학상, 노벨생리학상, 노벨평화상, 노벨화학상 등
 - 대통령상, 경제정의기업상, 자기상표제품 수출상, 대중상
- 본 문
 - 대중상은 한국영화계의 대표적 영화제로 존속하면서 우수영화의 발굴 및 포상을 통한 진흥효과를 거둔 것은 긍정적이지만
- 주의
 - '대상', '최우수상', '최고상', '우수상' 등의 일반 명사도 CV_PRIZE 에 해당됨. 그러나 '우승, 준우승'은 CV 에 해당하지 않음.
 - 상과 훈장의 명칭은 DATE, QUANTITY 와 함께 쓰이는 경우가 많다. 이 경우 DATE 와 QUANTITY 는 분리하여 태깅한다.
 - ◆ 2000 년 제 34 회 납세자의 날 대통령상
 - <DT_YEAR:2000 년> <QT_ORDER:제 34 회> <DATE:납세자의 날> <CV_PRIZE:대통령상>
 - ✕ <DT_YEAR:2000 년> <DATE:제 34 회 납세자의 날> <CV_PRIZE:대통령상>
 - ✕ <DT_YEAR:2000 년> <CV_PRIZE:제 34 회 납세자의 날 대통령상>
 - ◆ 제 10 회 경제정의기업상
 - <QT_ORDER:제 10 회> <CV_PRIZE:경제정의기업상>
 - ✕ <CV_PRIZE:제 10 회 경제정의기업상>

7.18 CV_LAW

- 정 의
 - 법/법률 명칭
 - 법적 권리
 - 범죄 명칭
 - 형벌 명칭
 - 칙령 등도 해당됨
- 법/법률
 - 예제
 - ◆ 국제연합헌장, 민법, 형법, 함무라비법전, 관습법 등
 - ◆ 헌법, 국회법, 법원조직법
 - 본 문
 - ◆ 실질적 의미의 헌법에는 형식적 의미의 헌법뿐 아니라 정부조직법·국회법·법원조직법·정당법·선거법 등의 법률과 ...
 - 주 의
 - ◆ '상고', '상소', '항소심' 등의 법률용어는 CV_LAW 가 아님
 - ◆ '대동법'은 조세 제도이므로 CV_POLICY 로도 볼 수 있지만, ~법으로 끝나는 경우에는 CV_LAW 로 태깅함
 - ◆ '방곡령' 등의 특정 정책의 명령은 CV_LAW 로 태깅함
 - ◆ '대사령' 등의 행정 명령은 CV_LAW 에 해당됨
- 법적 권리
 - 예 제
 - ◆ 피선거권, 참정권, 선거권, 공민권 등
 - ◆ 신분권, 친권, 부양청구권, 상속권, 어업권, 상표권, 저작권(소프트웨어 라이선스)
 - 본 문
 - ◆ 대의원 자격 요건은, 국회의원 피선거권이 있는 30 세 이상의 자로서 평화통일을 위해 주권을 행사할 수 있는 인물이어야 했다.
 - ◆ 법원의 판결 또는 다른 법률에 의하여 피선거권이 정지되거나 상실된 자는 피선거권이 없다
 - ◆ 선거일 현재 5 년 이상 국내에 거주하고 있는 40 세 이상의 국민은 대통령의 피선거권이 있다.
 - ◆ 참정권(參政權)도 광범위하게 보장되며
 - ◆ 1 회원이 가질 출자좌수는 총출자좌수의 100 분의 15 를 초과할 수 없다. 회원은 출자좌수에 관계없이 평등한 의결권과 선거권을 가진다
 - ◆ 수형(受刑) 중인 사람이나 선거법을 위반한 자에 대해서는 선거권·피선거권이 박탈되며, 이를 공민권의 정지·박탈이라고 한다.
 - ◆ 신분권에는 친족권과 상속권이 있는데, 친권, 후견권, 배우자에 대한 권리, 부양청구권 등은 친족권에 속하며, 재산상속권은 상속권에 속한다.
 - ◆ 어업권은 법적 성질이 물권임에도 불구하고 원칙적으로 이를 이전 ·분할 또는 변경할 수 없다.

- ◆ 상표권은 재산권의 일종으로서 특허권 등과 같이 담보에 제공될 수 있으며, 지정상품의 영업과 함께 이전할 수도 있다.

- 범죄 명칭

■ 예 제

- ◆ 간통, 살인죄 등
- ◆ 사체유기죄, 폭행죄, 공무집행방해죄, 명예훼손죄, 사기죄

■ 본 문

- ◆ 사체를 다른 곳으로 옮겨놓은 경우, 또는 그 죄적을 은폐하기 위하여 사체를 은닉한 경우에는 살인죄와 사체유기죄 또는 사체은닉죄와의 경합범(競合犯)이 된다.
- ◆ 폭행죄의 폭행은 사람의 신체에 대한 일체의 불법적인 유형력의 행사를 포함하게 된다.
- ◆ 공무집행방해죄는 공무를 집행하는 공무원에 대하여 폭행 또는 협박하여 직무집행을 하지 못하도록 하는 경우와
- ◆ 명예훼손의 행위(사실 또는 허위의 사실을 적시한 행위)와 모욕행위가 동시에 행해졌을 때에는 모욕행위는 명예훼손죄에 흡수되어 명예훼손죄만이 성립한다.
- ◆ 결혼을 가장하고 상대방을 기망하여 결혼과 패물비용 등의 명목으로 금품을 사취하는 이른바 혼인사기는 사기죄가 된다.

- 형벌 명칭

■ 예 제

- ◆ 화형, 태형, 구금 등
- ◆ 구류, 벌금, 몰수, 교수형, 총살형

■ 본 문

- ◆ 한국에서는 화형을 제도화하거나 관습화되었던 사례는 찾아볼 수 없으나
- ◆ 사형수는 모두 참형으로 집행하였다
- ◆ 구금이 비교적 장기에 걸치는 강제처분인 데 반해 구류는 가벼운 범죄에 대하여 형법상 다른 형과 함께 선택형으로 규정하고 있으며,
- ◆ 벌금은 판결 확정일로부터 30 일 이내에 납입하여야 하며
- ◆ 행위자에게 유죄의 재판을 하지 아니할 때에도 몰수의 요건이 있는 때에는 몰수만을 선고할 수 있다
- ◆ 한국에서는 원칙적으로 교수형이 사형집행 방법으로 되어 있고, 군인은 원칙적으로 총살형이 집행된다.

- 전문분야(법률)

- 탄핵, 계엄, 강화, 소추, 탄핵소추, 탄핵심판, 무죄판결, 긴급명령, 긴급재정경제처분 등과 같은 법률 용어는 개체명 대상이 아님
- 계엄령, 비상계엄, 경비계엄 등과 같이 실제 시행이 되는 정책 용어인 경우는 법률 용어이지만 CV_POLICY 로 태깅
- 국정감사, 불기소처분, 일사부재의는 CV_LAW 로 태깅

- 법률안, 예산안, 수정안은 특정한 법안이나 특정한 용어를 가리키는 것이 아니기 때문에 개체명 대상이 아님
- 처벌(형벌) 및 법률 관련
 - 벌금 200 만원, 징역 12 개월과 같은 처벌(형벌)은 각각 분리하여 태깅한다. 단, 벌금 OO 형, 징역 OO 형, 집행유예 OO 형은 통합하여 태깅한다.
 - ◆ 벌금 200 만원 : 벌금/CV + 200 만원/QT 로 태깅한다.
 - ◆ 징역 12 개월 : 징역/CV + 12 개월/DT 로 태깅한다.
 - ◆ 징역 1 년, 집유 6 월 : 징역/CV + 1 년/DT , 집유/CV + 6 월/DT
 - ◆ 징역 3 년형, 집행유예 1 년형 : 징역 3 년형/CV, 집행유예 1 년형/CV
 - 법률 조항의 경우도 각각 분리하여 태깅한다.
 - ◆ OO 법 제 3 조 제 4 항 : OO 법/CV + 제 3 조/QT + 제 4 항/QT
 - '벌금'은 법적 효력이 있는(벌금형 선고) 경우에만 태깅하며, 한 사회나 공동체에서 규칙으로 정한 벌금의 경우에는 태깅하지 않는다.

7.19 CV_FOOD_STYLE

- 정 의
 - 음식 종류
- 예 제
 - 한식, 일식, 양식, 중식
 - 보양식, 패스트푸드, 육식, 채식
 - 한국음식, 중국음식, 일본음식, 미국음식
- 본 문
 - 오, 1 등석은 디저트도 나오는구나.
 - 흥합은 고단백 저지방 식품이면서 ...
 - 콜로이드와 단백질이 풍부하여 여름보양식으로 유명한 삭스핀을 먹었다.
- 주 의
 - '절식'은 음식 종류가 아니므로 CV_FOOD_STYLE 에 해당되지 않음

8. DATE

DATE 는 날짜 표현을 나타낸다.

8.1 DT_DURATION

- 정 의
 - 기간
- 예 제
 - 약 5500 만~3000 만 년 전, 약 40 년 동안, BC 624~BC 544 년, 10 년간, 6~7 월
 - ~부터 ~까지, 일년 내내, ~ 간, 전반기/후반기, 성수기/비수기
 - 하루
- 본 문
 - 우리나라 한 해 수출규모의 16 분의 1 수준인데요.
 - 과거 50 년 동안 한국이 가난에서 벗어나서 경제발전을 했던 그 경험들, 그러한 것들을 OECD 가입자들에게 접목을 시키고.
- 주 의
 - "봄철 해빙기"에서 '해빙기'의 경우는 DT_DURATION 대상이 아님
 - "프랑스 혁명 시기" 등의 ~ 시기는 DT_DURATION 으로 태깅함
 - '유년시대'는 DT_DURATION 대상이 아님
 - 정확한 기간을 나타내는 경우에만 태깅한다.
 - ◆ '~부터 ~까지, ~간'은 명확한 시간, 기간을 의미할 때 태깅한다. 예) 2006 년부터
 - ◆ '겨울방학 기간' 역시 정확한 기간을 의미하지 않으므로 태깅하지 않는다.
 - '월 3 만 9 천원'에서 '월'은 태깅하지 않는다.

8.2 DT_DAY

- 정 의
 - 특정 일/절기
- 예 제
 - 21 일, 20 일
 - 입춘, 곡우, 첫날, 첫주, 마지막 날, 내일, 모레, 금일, 당일, 하지
- 본 문
 - 오늘 오후면 서울 G20 정상회의는 마무리됩니다.
 - 지난 4 일 프랑스와 중국간의 정상회담이 열렸습니다.
 - 20 일 경기

8.3 DT_MONTH

- 정 의
 - 달
- 예 제
 - 9 월, 8 월
 - 이달, 저번달, 올 4 월
- 본 문
 - 지난 8 월에 내년도 파리회의에서 사무국 설치문제를 의제화하겠다고 프랑스가 발표했습니다.
 - 현장에서 들려오는 소식으로는 지난 달 경주에서 열렸던 ...
 - 제가 9 월에 발표했었던 내용이 다시 언급되기도 했었었는데요.
- 주 의
 - '매일', '매월'은 태깅하지 않는다.

8.4 DT_YEAR

- 정 의
 - 년
- 예 제
 - BC 3000 년, BC 300 년경, 880 년, 1970 년대
 - 작년, 내년, 올해, 서기 1910 년, 영락 1 년
- 본 문
 - 이번 회의에서 결론을 내지 못하고 내년 프랑스 정상회의로 미루었습니다.
 - 예를 들어서 관광산업을 보면 환율이 제일 약했고 말하자면 외국 사람들이 제일 유리했던 것은 2008 년때, 그때 달러의 환율이 1410 원까지 올라갔었어요.
- 주 의
 - '영락'과 같이 연호만 있는 경우는 개체명(DT_YEAR) 대상이 아님
 - '~년경'의 경우도 DT_YEAR 임

8.5 DT_SEASON

- 정 의
 - 계절
- 예 제
 - 봄, 여름, 가을, 겨울
 - 작년 봄, 올 가을, ~시즌(예: 2008 년 시즌, 스토브리그, 포스트시즌)
- 본 문
 - 올 봄 유럽을 뒤흔친 그리스발 금융위기.
 - 11 월의 날씨답지 않게 초겨울 날씨를 보이고 있는 요즘인데요.

8.6 DT_GEOAGE

- 정 의
 - 지질시대
- 예 제
 - 캄브리아기, 중생대 초기, 중생대, 원생대, 올리고세, 에오세, 신생대 제 3 기, 신생대, 선사시대, 석탄기, 빙기, 고생대 오르도비스기, 고생대
 - 청동기시대, 구석기시대, 신석기시대
- 본 문
 - 캄브리아기의 지층으로부터는 유공충 ·방산충(放散蟲:原生動物) ·해면(해면동물) ·해파리(강장동물) ·바다나리(극피동물) ·완족류[擬軟體動物] ·고둥 ·앵무조개(연체동물) ·삼엽충(三葉蟲:절지동물) ·필석류(筆石類:원색동물) 등의 동물화석이 다량으로 발견된다.

8.7 DT_DYNASTY

- 정 의
 - 왕조/시대
- 예 제
 - 청대, 조선시대, 조선 후기, 조선 중기, 명대, 고려시대, 고려 말, 고대 로마제국, 고구려시대, 가리에누스, 황제 시대, 5 현제시대, 당대, 고려시대, 조선 정조 23 년, 영조 27
 - 중세
- 본 문
 - 조선시대 궁중무용인 정제는 음악과 노래와 춤이 삼위 일체를 이루는 귀한 예술이죠.
 - 대학 공사현장에서 구한말과 일제강점기 것으로 추정되는 문화재가 발굴됐습니다.
 - 고대 신라사를 새롭게 쓰게 했죠.
- 주 의
 - "구한말", "일제강점기", "조선 시기" 등은 DT_DURATION 으로 볼 수도 있지만, 역사의 특정 시대를 나타내므로 DT_DYNASTY 로 태깅한다.
 - 왕의 이름 뒤에 "~때"가 붙으면 DT_DYNASTY 로 태깅
 - ◆ 성종 때, 연산군 때 등
 - '현대'는 '중세'와 '고대'와 달리 시기를 확정할 수 없으므로 일단 태깅 대상에서 제외한다.
 - 왕의 이름 뒤에 "~때"가 붙으면 DT_DYNASTY 로 태깅
 - ◆ 성종 때, 연산군 때 등
 - '현대'는 '중세'와 '고대'와 달리 시기를 확정할 수 없으므로 일단 태깅 대상에서 제외한다.

8.8 DT_OTHERS

- 정 의
 - 기타 날짜
 - DATE 의 세부분류에 해당되지 않는 날짜
- 예 제
 - ~년 후/전, ~부터/까지/정도, 지난 세기, ~ 때
 - 제 2 차 세계대전 전, 임진왜란 때
- 본 문
 - 대릉원은 4 세기에서부터 6 세기경 신라 지배계층의 무덤이 모여 있는 곳입니다.
 - 1981 년부터 매년 1 월에서 2 월 사이 휴양 도시 다보스에서 열렸다.
- 주 의
 - "년/월/일, 년/월, 월/일"이 함께 쓰일 경우에는 전체를 묶어서 DT_OTHERS 로 태깅한다.
 - ◆ <2007 년 10 월:DT_OTHERS>에 체급을 올렸다.
 - ◆ <작년 11 월 25 일:DT_OTHERS> OECD 에 들어갔다.
 - "예로부터"는 DT_OTHERS 대상이 아님
 - 오늘부터, 작년 봄부터, 올해부터, 가을 시즌부터, 포스트 시즌부터와 같이 '부터'를 제외했을 때에도 개체명으로 태깅할 수 있는 것은 모두 DT 로 태깅한다.
예) 오늘부터/DT
 - 그러나 '이제부터', '지금부터'와 같이 '이제'와 '지금'과 같이 '부터'를 제외했을 때 개체명으로 태깅할 수 없는 불명확한 시기를 나타내는 것에 대해서는 태깅하지 않는다.
 - '납세자의 날'을 'DT'로 태깅하기 때문에 설날, 크리스마스 등 주기적으로 나타나는 날이나 특정일로 판단되는 것은 DT 로 태깅한다.
 - '매해마다', '매일', '매월', '매주', '매년'은 태깅하지 않는다.
 - '지난'은 단독 태깅하지 않으며, '지난+DT 범주에 속하는 명사 또는 명사구'일 경우에만 태깅한다.
 - ◆ 지난 아시안게임 : 아시안게임/EV 만 태깅, 지난 정부 : 정부/OG 만 태깅
 - '작년 인천아시안게임 때', '지난 2002 년 한일 월드컵 때'와 같이 'DT+EV+~때'의 형태의 경우 통합하여 태깅하지 않고 각각 따로 태깅한다. ('EV+~ 때'는 '임진왜란 때/DT'처럼 합쳐서 DT 로 태깅)
 - ◆ 작년 인천아시안게임 때 : 작년/DT + 인천아시안게임 때/DT
 - ◆ 지난 2002 년 한일 월드컵 때 : 지난 2002 년/DT + 한일 월드컵 때/DT
 - '~ 전, 후, 이전, 이후, 직후, 직전' 등은 EV, DT, TI 에 한정해서 태깅한다.

9. TIME

TIME 은 시간 표현을 나타낸다.

9.1 TI_DURATION

- 정 의
 - 기간
- 예 제
 - 6 시~9 시, 오후 1 시~3 시, 10~15 분
 - ~ 시간, ~시간 동안
 - 아침, 낮, 밤
- 본 문
 - 이들은 피해자를 다섯 시간 넘게 끌고 다녔다.
 - 내일은 서울의 낮 기온이 11 도까지 오르면서 예년 이맘때 기온을 되찾겠습니다.
 - 전국에 가끔 구름이 지나겠고 아침에는 내륙지역에 짙은 안개 끼는 곳이 있겠습니다.
 - 오전 10~11 시경부터 높아져서 한낮이 지날 무렵 최고가 된다.
- 주 의
 - 기간이나 지속 시간 등은 모두 TI_DURATION 에 해당됨
 - 스포츠 경기의 시간 기록은 TI_DURATION 에 해당됨 (예: 올림픽경기대회에서도 2 시간 12 분 11 초라는 당시 세계최고기록을 이룩하였다)

9.2 TI_HOUR

- 정 의
 - 시간
- 예 제
 - 12 시, 저녁 7 시
- 본 문
 - 마르코스 씨는 6 시쯤에 집에 도착했습니다.
 - 수험생들은 오늘 오전 재학중인 학교 등에서 시험장이 기재된 수험표를 배부받은 뒤 오후 3 시 시험장으로 찾아가서 예보소집에 참가해야 합니다.
- 주 의
 - 특정 시각만 TI_HOUR 에 해당됨
 - 소요되는 시간은 TI_DURATION 으로 태깅함 (예:서버를 34 시간 운영하고 있다).

9.3 TI_MINUTE

- 정 의
 - 분
- 본 문
 - 지금 시간이 2010 년 마지막 11 시 56 분, 57 분을 지나고 있습니다.
 - 갓낚은 새끼는 연한 노란색으로 생후 20 분이면 달릴 수 있다.
 - 김해국제공항까지의 거리가 약 20 분 단축되었다.
- 주 의
 - 특정 시각의 분이나 지속 시간 이외의 분 단위 시간을 말할 때만 TI_MINUTE 로 태깅함

9.4 TI_SECOND

- 정 의
 - 초
- 본 문
 - 300m 를 주파하는 데 걸린 시간은 단 15 초입니다.
 - 평균 액세스타임도 0.5 초 정도 빠르다.

9.5 TI_OTHERS

- 정 의
 - 기타 시간
 - TIME 의 세부분류가 아닌 기타 시간
 - ~시간 전/후
- 본 문
 - 지금 시간이 2010 년 마지막 11 시 56 분, 57 분을 지나고 있습니다.
 - 러시아에서 발사되었으며 6 시간 후 지상국과 교신에 성공하였다.
- 주 의
 - 3 회 말, 9 회 초는 문맥에 따라 **TI** 로 태깅한다.
 - '전, 후, 이전, 이후, 직후, 직전 등'은 EV, DT, TI 에 한정해서 태깅한다.

10. QUANTITY

수량 표현을 나타낸다.

10.1 QT_AGE

- 정 의
 - 나이
- 예 제
 - 만 55 세, 50~60 대, 45 세, 3 살, 18~41 세
- 본 문
 - 어머니는 선지인데, 인수 4 만 세 때에 안화성에서 태어나 시리수 아래서 성불하였다.
 - 어느 광산촌에서 아버지마저 잃고 가난하게 살아가던 10 세 소녀 안말숙 4 남매는 뽀뽀이 헤어지는 슬픔을 겪는다.

10.2 QT_SIZE

- 정 의
 - 크기/넓이/면적
- 예 제
 - 9,606.33 km², 6 만 8,000 km², 57.7×138.5cm, 430×230 m, 3 만 1500 km², 30 만 정보, 28.02 km², 20 만 정보, 20 km², 2.09 km², 2,448 km², 2,200 km², 19×7.5 m, 192 만 2570 km², 173 만 km²
- 본 문
 - 면적은 33.3 km², 인구는 19 만 3214 명이다.
 - 보통 1 제곱미터의 땅, 그러니까 이 정도 넓이의 땅에는 200 마리 정도의 지렁이가 있는데요.

10.3 QT_LENGTH

- 정 의
 - 길이/거리/높이
- 예 제
 - 두께 0.76mm, 가로 54mm, 900m, 8km, 82 피트, 7cm, 6~12cm, 3~5 cm, 2,200km,
- 본 문
 - 가항(可航)거리는 상류에서 700km, 하류에서 300km이며, 종류는 항행이 불가능하다.
 - 르커저시[日喀則市]와의 거리는 320km 이다.
 - 1852년에는 두 핵 사이의 거리가 2.5×106km 인 쌍혜성이 되어 출현하였는데
 - 높이 4,182m 이다.
 - J 자형의 섬으로 높이 2,987m 에 달하는 산들이 솟아 있으며
 - 해발고도 3,104m 의 고지에 자리잡고 있다.

10.4 QT_COUNT

- 정 의
 - 개수, 빈도
- 예 제
 - 6 개, 8 백 70 선, 하나, 둘
 - 50 개국, 1 마리, 약 600 종, 약 3 만 종, 약 1,000 종, 9~10 마리, 14 종, 10 권, 두 차례, 9 개, 97 건, 7 천 명~2 만 명, 7 가지, 77 만 9472 건, 6 가지, 50 개국, 5,000 자, 4 차례, 45 만 명, 4,123 석, 390 여 칸, 18 행, 13 척, 1,600 여 개
- 본 문
 - 거머리들의 몸은 정확히 34 개의 환절로 이루어져 있는데 그 이상도 그 이하도 아니죠.
 - 어떤 거머리들의 입은 이렇게 세 갈래로 갈라지는데 한쪽마다 100 개도 넘는 이빨이 나 있어서 마치 톱처럼 먹이의 피부를 자를 수 있죠.
- 주 의
 - “오류”과 같은 경우는 대상이 아님
 - 질량수는 QT_COUNT 임
 - ‘수 관형사 + 단위명사, 수 관형사 + 일반명사’는 모두 통합하여 태깅하는 것으로 수정. 일반명사가 개체명이든 비개체명이든 모두 통합하여 태깅한다.
 - ◆ 두 마리 : 두 마리/QT 두 감독 : 두 감독/QT, 두 여자 : 두 여자/QT

10.5 QT_MAN_COUNT

- 정 의
 - 인원수
- 예 제
 - 50 명, 1 명, 약 600 명, 12 인, 2 억 1219 만 5000 명
- 본 문
 - 경찰은 주지훈 씨 외에도 10 여 명의 마약투약혐의를 포착하고 혐의를 입증하는 데 주력하고 있어 연예계 마약파문은 당분간 계속 이어질 것 같습니다.
 - 많은 스태프들에게 둘러싸여 충실력을 뽐내고 있는 6 명의 스타들.
- 주 의
 - 명확히 인원수를 나타내는 경우에만 태깅한다.

10.6 QT_WEIGHT

- 정 의
 - 무게
- 예 제
 - 30 kg, 140g, 백근, 9 만 t, 85~100 kg, 60g, 5 만 톤, 170t
- 본 문
 - 키 1m 80cm 의 노모스카는 몸무게가 고작 49kg.
 - 여기서 가장 큰 타이어인데 지름이 4m 이고 무게가 5 톤이래요.

10.7 QT_PERCENTAGE

- 정 의
 - 백분율, 비율, 농도
- 예 제
 - 약 75%, 90%, 8~12ppm, 80~90%, 73.7%, 4 배, 400~500ppm, 40%, 3 배, 3~5ppm, 32.3%, 30~40%, 2 배, 26~34%, 20%, 16%, 11~32%, 10~20ppm, 10ppm, 100ppm, 1/3, 1.1%, 0.719 %, 0.1ppm, 1/4, 4 분의 1
- 본 문
 - 자동차 타이어에는 60% 정도의 합성고무가 들어가고 트럭 타이어에는 이보다 약간 덜 들어간대요.
 - 보통 1 년에 공기압력이 4 분의 1 정도 줄어든조.

10.8 QT_SPEED

- 정 의
 - 속도
- 예 제
 - 1km/s, 30km/s
- 본 문
 - 시속 100km 까지 달리다가 브레이크를 밟는 거군요.
 - 이들 중성자는 믿기지 않을 정도의 엄청난 속도로 무려 초속 2 만킬로미터로 피복을 뚫고 나오면 물 때문에 속도가 느려지면서 옆의 연료봉으로 뚫고 들어가죠.

10.9 QT_TEMPERATURE

- 정 의
 - 온도
- 예 제
 - 22~25°C, - 2.0°C, 9°C~12°C, 41~43.5°
- 본 문
 - 핵분열로 뜨거워진 물은 원자로 내부에서 320 도까지 올라갑니다.
 - 800 도부터는 연료를 감싸고 있는 피복에 금이 가면서 방사선물질이 갈라진 틈 사이로 빠져나갑니다.

10.10 QT_VOLUME

- 정 의
 - 부피
- 예 제
 - 두 말, 800~1,000 mm, 500mm, 2 만 2000 km³, 20ℓ, 2,000~2,500mm, 0.15µg/ℓ
- 본 문
 - 올리브유 1 리터를 얻으려면 적어도 올리브 5kg 은 따야 해요.
 - 이곳에서는 1 세제곱미터당 먼지입자를 33 개 미만으로 유지해야 하는데 1 세제곱미터는 보통 냉장고만 한 공간이에요.

10.11 QT_ORDER

- 정 의
 - 순서/순차적 표현
- 예 제
 - 제 82 호, 제 7 권, 6 기, 65 위, 제 9 회, 제 4 교구, 제 2 차, 제 2 대, 제 2B 족, 제 1 대, 제 117 호, 7 층, 5 차, 4 차, 3 층, 12 대
 - 1 위, 재수, 3 수
 - 4 옥타브
- 본 문
 - 그런데 제이미, 이게 아직 1 단계에 불과하다네요.
 - 7 위는 무한도전 듀엣가요제.
- 주 의
 - 원자번호는 QT_ORDER 임
 - 1 쇠, 2 쇠, 1 권, 2 판 등은 QT 로 태깅한다. 단, 증보판, 재판은 태깅하지 않는다.
 - '첫'은 관형사이므로, 단독으로는 태깅하지 않는다.
 - '첫 + 일반 명사'는 일반명사가 DT(날짜 관련 일반 명사)일 때만 전체를 DT 로 태깅한다.
 - ◆ 첫+ 날짜 관련 일반 명사 : 첫날, 첫 주, 첫 달 → DT 로 태깅

- '첫 번째', '첫 장'은 QT 로 태깅한다. ('장' 명사)
- 전문분야(특허)
 - 특허분야는 이전의 도면과 청구항을 나타내는 패턴이 존재
 - 특정한 기능, 장치 등의 뒤에 괄호로 번호가 명시되는 경우가 존재
 - ◆ 도 1, 도 2,..., 청구항 1, 청구항 2,...
 - ◆ ~장치(100), ~모듈(200) -> 괄호안의 숫자는 QT_ORDER
 - ◆ S10, S30, S40, S55, S60, ...

10.12 QT_PRICE

- 정 의
 - 금액
- 예 제
 - 72 만 원, 60 억 달러, 40 억 달러, 403 원, 290 원, 1 조 2 천억 원, 188 원, 100 만원, 100 만 원
- 본 문
 - 소더비 경매시장에서 무려 200 만달러, 약 26 억원 되는 돈에 낙찰됐습니다.
 - 윤상현 씨가 전 소속사로부터 10 억원대의 손해배상 소송에 휩싸였습니다.

10.13 QT_PHONE

- 정 의
 - 전화번호
- 예 제
 - 642-101-1010, (T)<QT_PHONE: 011-111-0111> 등
- 본 문
 - 이들 행사의 참가비는 없으며, 선착순 입장해야 한다. (문의: ☎345-6313)
 - 긴급통신(119 · 112) 및 감청 방식이 가닥이 잡혔다.

10.14 QT_SPORTS

- 정 의
 - 스포츠 관련 수량 표현
- 예 제
 - 1 승 2 패, 5 타점, 3:2, 2 연승
 - ~경기, ~승/패/무, 10 호골, ~안타
- 본 문
 - 아스날은 이번 경기에서의 승리로 칼링컵 포함 14 경기에서 13 승 1 무를 하며 무패 행진을 이어가게 되었다.

- 아데바요르는 왼쪽으로 가볍게 성공시키며 1:0 으로 앞서게 되었고 그 골을 그대로 지켜내며 결승골로 만들어 1:0 승리를 거두었다.
- 주 의
 - 볼넷, 삼진과 같은 스포츠 규칙 관련 용어나 수량의 의미가 없는 1, 2 루와 1, 2 레인은 TM_SPORTS 로 태깅한다.
 - 6 이닝 은 QT 로 태깅한다. '1 볼넷, 6 오버파'의 경우 QT 로 태깅한다.
 - 3 회 말, 9 회 초는 **TI** 로 태깅한다.

10.15 QT_CHANNEL

- 정 의
 - TV/라디오 채널 번호
- 예 제
 - 44 번 채널, 6 번, 20 번, 91.9, 95.9 MHz, 39 쇼핑,
- 본 문
 - tvN 의 채널번호가 17 번에서 3 번으로 변경됩니다.
- 주 의
 - 명확한 채널 번호에 해당되는 경우에만 태깅한다.
 - CJ 홈쇼핑, KBS 2TV 의 경우는 채널명에 해당되기 때문에 AF(프로그램명)으로 태깅한다. 예) CJ 홈쇼핑/AF , 코미디 TV/AF , MBCsports+/AF
 - 문맥에서 채널명이 아닌 방송 기관으로 판단될 경우 AF 가 아닌 기관명에 해당하는 OG 로 태깅할 수 있다.

10.16 QT_ALBUM

- 정 의
 - 앨범 관련 수량 표현
- 예 제
 - 1 집, 2 집 등
- 본 문
 - 5 집 발매 보아, '남성 우월주의 거부한다'
 - 작곡가 최희주와 작사가 고은주가 주축이 돼 이끌고 있는 프로젝트 그룹 DnD 가 여름을 맞아 싱글 앨범 2.5 집을 발매했다.

10.17 QT_ZIPCODE

- 정 의
 - 우편번호
- 예 제
 - 110-704, 200-776
- 본 문
 - 760-905 경북 안동시 태화동 677-6.
 - 이마트 쌍용점 인근 서북구 서부대로 319-4 로 이전한다고 17 일 밝혔다.

10.18 QT_OTHERS

- 정 의
 - 기타 수량표현
 - QUANTITY 의 세부 분류가 아닌 기타 수량
- 예 제
 - 지식머니 20, 북위 38° 53' 51
- 본 문
 - 지식 머니는 얼마나 걸어야 할지는 모르겠지만 자세히 적어주시면 10 지식머니를 조금 어설프다 싶으면 5 지식머니를 드릴게요.
 - 한국에서는 0532 가 아니고 0 을 빼고 532 입니다.
- 주 의
 - '첫 여행', '첫날', '첫 주말', '첫 출발' 등은 QT_OTHERS 의 대상이 아님. '첫 + 날짜 관련 일반 명사'인 '첫날, 첫 주말'은 DT 로 태깅하며, 그 외의 '첫 + 일반 명사'는 태깅 대상이 아님.

11. EVENT

EVENT 는 특정 사건/사고 명칭을 나타낸다.

11.1 EV_ACTIVITY

- 정 의
 - 사회운동 및 선언
- 예 제
 - 공민권운동, 노예해방선언 등
 - 인권선언, 브나로드운동, 새마을운동, 양무운동
- 본 문
 - 1962 년 흑인의 투표권등록운동이 시작되었다
 - 인권선언은 1891 년의 프랑스 헌법을 비롯하여 그 후 세계 여러 나라의 헌법 및 정치에 크게 영향을 미쳤다.
 - 1931 년부터 1934 년까지 동아일보사가 전개한 브나로드운동은 고등보통학교 4, 5 학년 학생으로 이루어진 학생계몽대와 전문학교 이상의 학생으로 조직된 학생강연대, 학생기자대를 주축으로 하여 행해졌다.
 - 새마을운동은 1970 년대의 한국사회를 특징짓는 중요한 사건이다.
 - 그러나 이 양무운동이 중국 자본주의 발달의 싹이 되었다는 것은 부인할 수 없다.
- 주 의
 - “~운동”, “~선언” 등이 해당됨
 - “갑오개혁”, “갑오경장”도 해당됨
 - 독립운동에 해당하는 사건도 EV_ACTIVITY 에 해당됨
 - “~칙령”은 CV_LAW 에 해당됨
 - 사보타주는 노동운동보다는 특정 행동으로 볼 수 있으므로 EV_ACTIVITY 의 대상이 아님

11.2 EV_WAR_REVOLUTION

- 정 의
 - 전쟁/혁명
 - 전투
- 예 제
 - 신해혁명, 포클랜드전쟁, 아편전쟁, 메이지유신
 - 러일전쟁, 청교도혁명, 프롤레타리아혁명, 4.19 혁명, 남북전쟁

- 본 문
 - 러·일전쟁의 결과는 포츠머스 강화회담과 을사조약으로 이어져
 - 프롤레타리아혁명은 1917 년 러시아에서 레닌의 지도하에 실현되었으며, 그 후 동유럽과 중국·쿠바·베트남 등으로 확산되어 갔다.
 - 4·19 혁명은 한국의 정치발전사에 하나의 획기적인 전기(轉機)를 기록한 역사적으로 큰 의미를 지니는 일대사건(一大事件)이었다.
- 주 의
 - ~전쟁/~혁명을 대상으로 함
 - '노르망대상륙작전' 등의 전투작전도 포함됨
 - '세계 2 차대전' 등의 전쟁명칭도 포함됨
 - '메이지유신'은 EV_OTHERS 에 해당됨
 - 무력 정변은 EV_WAR_REVOLUTION 에 해당됨 (예: 갑신정변, 갑오개혁)
 - 민란, 반란, 반정은 EV_WAR_REVOLUTION 에 해당됨(예: 임술민란, 이인좌의 난, 인조반정)
 - 농민봉기는 EV_WAR_REVOLUTION 에 해당됨 (예: 임술농민봉기)
 - 모반 사건은 EV_WAR_REVOLUTION 에 해당됨 (예: 정여립 모반 사건)

11.3 EV_SPORTS

- 정 의
 - 스포츠 관련 행사
 - 게임 관련 e 스포츠 이벤트(행사) 이름
 - 특정 고유행사명이 아니라, 일반 유형(예: 마라톤 대회)의 경우도 포함
- 예 제
 - 올림픽경기대회, 아시아경기대회, 서울 올림픽 경기, 88 서울올림픽, 서울올림픽
 - ~대회/경기(예: 거제시장기 전국 윈드서핑대회, 홈경기, 원정경기), ~전(예: 두산전, 개막전, 최종전, 결승전, 예선전, 주중 3 연전)
- 본 문
 - 특히 올해 열세 번째를 맞는 수중 마라톤 대회는 황강의 물줄기를 이용한 독특한 코스를 달리는 전국 유일의 마라톤 대회로 마라톤 마니아뿐 아니라 가족 단위의 참가자들에게도 인기를 끌고 있다.
 - 해발 1 천 567m 태백산 정상에 올라 눈 꽃을 감상하는 등반대회를 끝으로 막을 내리게 된다.

- 주 의

- 스포츠 행사는 보통 개최 회수, 지역 등과 같이 쓰인다.
- 개최 회수는 분리하여 태깅하고 지역은 통합하여 태깅한다.
 - <QT_ORDER:제 24 회> <EV_SPORTS:서울 올림픽 경기>
 - ✕ <QT_ORDER:제 24 회> <LCP_CAPITALCITY:서울> <EV_SPORTS:올림픽 경기>
 - ✕ <QT_ORDER:제 24 회 서울 올림픽 경기>
- 스포츠 행사가 DATE 와 병행하여 쓰일 경우 명확한 연도 표시는 독립하여 태깅하고, 축약된 경우는 통합하여 태깅한다.
 - <DT_YEAR:1988 년> <EV_SPORTS:서울 올림픽 경기>
 - <EV_SPORTS:88 서울 올림픽 경기>
 - ✕ <EV_SPORTS:1988 년 서울 올림픽 경기>
 - ✕ <DT_YEAR:88> <EV_SPORTS:서울 올림픽 경기>
- 스포츠 행사의 경우 '4 강, 8 강' 등과 같이 숫자가 나와도 문맥에 따라 QT 가 아닌 EV 로 처리해야 하는 경우가 있으므로 주의해야 한다.
 - ◆ '4 강, 8 강' 등 앞에 숫자가 나온 경우, '4 강전'과 같이 '~전'을 의미하면 EV 로, 4 개의 강팀을 의미하면 QT 로 태깅한다. (프로야구는 6 강전, 5 강전을 진행함)
 - ◆ '2 라운드'와 같은 표현이 경기의 수를 세는 의미라면 QT 로, 예선전, 결승전과 같이 경기 자체를 의미하는 것이라면 EV 로 태깅한다. (예: 이번 K 리그 축구는 총 32 라운드를 경기할 예정이다. → 32 라운드/QT)
 - ◆ '~회, ~세트' 역시 이에 준하여 태깅한다. (예: '대역전극은 9 회에서 펼쳐졌다.', '역전 드라마는 마지막 5 세트에서 펼쳐졌다' → 9 회/EV, 5 세트/EV)
 - ◆ '3 파전'은 EV 로 태깅한다.
 - ◆ 3 회 말, 9 회 초는 TI 로만 태깅한다.
 - ◆ 정리
 - 가. 라운드 → 1 라운드/QT or 1 라운드/EV
 - 나. 1 세트 → 1 세트/QT or 1 세트/EV
 - 다. 예선 1 경기 → 예선 1 경기/QT or 예선 1 경기/EV

11.4 EV_FESTIVAL

- 정 의

- 축제 명칭

- 예 제

- 연어축제, 갯방구대축제, 새미골막사발축제 등
- 연등축제, 이천도자기축제

- 본 문

- 1973 년 시작된 제등행진이 1996 년 새롭게 단장되어 문화축제로 발전된 것이 연등 축제이다.
- 이천도자기축제는 우수한 이천도자기를 전세계에 널리 알리고, 우리 문화의 저변 확대와 지역 활성화를 위해 이천문화원 이천민속도자기사업협동조합 주관

- 주 의
 - "~축제", "~영화제"만 대상으로 함

11.5 EV_OTHERS

- 정 의
 - 기타 사건/사고 명칭
 - ~사태 (ex. 페르시아만사태, 한진중공업파업사태, 중동사태, 천안문사태)
- 예 제
 - 태안 기름유출사고, 북경세계대회
 - 축제 내 체험행사 (예: 바다낚시 체험)
- 본 문
 - 대형 쥐 문양판에 끼우는 무자년 소원 담기 행사를 비롯해 옷을 던져 새해 운수를 점치는 옷점 보기, 복 쥐띠 문양 페이스페인팅 프로그램 등이 진행된다.
 - 희망찬 출발을 기원하는 내용의 비나리 공연을 펼친다.
- 주 의
 - EV_ACTIVITY와 EV_WAR_REVOLUTION은 구체적인 사건에 한정하여 태깅함.
 - EV_SPORTS와 EV_FESTIVAL은 일반 유형도 태깅함.
 - EV_OTHERS는 의미 범주에 따라 처리가 다름. '정상회담, ~투표, ~ 선거, ~시험/고사'의 경우 '선거, 투표, 시험'은 태깅하지 않으나 구체적인 날짜를 상정할 수 있는 사건이 아니더라도 '주민 투표, 주지사 선거, 중간 선거', '수능, 임용고사' 등은 태깅한다. 그러나 이 부류를 제외하면 구체적인 시기나 연도를 상정할 수 있는 사건, 사고, 사회적 현상에 한정하여 태깅한다.
 - 특정 미술 전시회 등의 전시회도 EV_OTHERS에 해당됨
 - "헤이그특사 사건", "헤이그밀사 사건" 등은 EV_OTHERS로 태깅함
 - '주민투표제'는 CV_POLICY에 해당되지만, '주민 투표'는 EV_OTHERS에 해당됨
 - TED와 같은 강연은 전시회와 같은 개념으로 EV_OTHERS로 태깅
 - 태풍 등 자연재해의 명칭은 "태풍 루사"와 같이 태풍과 이름을 모두 합쳐서 하나의 사건으로 보고 EV_OTHERS로 태깅. 향후 태그 확장시 세부태그 지정 필요.
 - "향토예비군훈련"과 같은 군사 훈련 역시 EV_OTHERS로 태깅.
 - EV는 내부에 다른 부류의 개체명이 포함되어 있어도, 쪼개지 않고 통합하여 하나로 처리한다.
 - ◆ '남북정상회담/EV', '한미정상회담/EV', '한일월드컵/EV', '완두콩 따기체험/EV'
 - '~정상회담'이나 '~협상'도 EV로 처리한다. '정상회담'은 예외적으로 구체화되지 않고 '정상회담'이 단독으로 나왔을 경우에도 태깅한다.
 - ◆ 중동평화협상/EV
 - ◆ 정상회담/EV, 한일정상회담/EV

12. ANIMAL

생물은 2004 년 스미스 분류체계에 따라서 다음과 같이 분류된다.

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

ANIMAL 은 진핵생물 중에서 동물에 해당되는 것으로, 사람을 제외한 짐승을 대상으로 한다.
세분류에 포함되지 않는 동물명은 AM_OTHERS 로 태깅한다.

12.1 AM_INSECT

- 정 의
 - 곤충
- 예 제
 - 버메뚜기, 가시거미, 지네, 불나방, 배추흰나비, 참매미, 사마귀
 - 타란툴라
- 본 문
 - 딱정벌레는 어른 벌레 상태로 영하 60 도까지 견딘다.
 - 특히 호주산 나비들만 채집되어 있는 나비관(Butterfly House)은 멜번 동물원의 자랑거리이다.

12.2 AM_BIRD

- 정 의
 - 새(조류) 이름
- 예 제
 - 왜가리, 삿꾸기, 거위, 두루미, 황새, 봉황
- 본 문
 - 거위는 가금이 된 지 오래되었으므로 거친 먹이도 잘 먹고 추위에도 강하다.
 - 실제로 이 그림은 두루미[鶴]가 아니라 황새[鷺]가 잘못 그려진 것이라 한다.
- 주 의
 - '덧새' 등의 새 종류는 AM_BIRD 에 해당하지 않음
 - '새'라는 단어는 AM_TYPE 에 해당하는 것으로 보고 태깅함.

12.3 AM_FISH

- 정 의
 - 물고기(어류) 이름
- 예 제
 - 참가자미, 전기뱀장어, 연어, 갈치, 멸치, 미꾸리, 송어, 날치
- 본 문
 - 《난호어목지》에서는 갈치(葛侈)라 하였다
 - 옛부터 식용과 약용으로 널리 쓰인 미꾸리는 지금도 맑은 물에 담궈 놓아 오물을 모두 토하게 한 후, 폭 고아서 국을 끓인 추어탕의 원재료이다.
 - 송어는 예로부터 음식으로서만 아니라 약재로도 귀하게 여겼다.
 - 날치 무리 중에는 열대성 날치와 아열대성 날치, 온대성 날치 등이 있다.

12.4 AM_MAMMALIA

- 정 의
 - 포유류
- 예 제
 - 해달, 토끼, 오소리, 돌고래, 고래, 박쥐, 고슴도치, 두더쥐, 당나귀
- 본 문
 - 일반적으로 몸길이 4~5m 이상인 것을 고래, 소형인 것을 돌고래라고 하는데, 망치고래와 같이 소형인 것도 고래라는 이름이 붙은 종류가 있다.
 - 겨울이 되면 일부 이주하는 박쥐를 제외하고 대부분이 겨울잠을 잔다.
 - 당나귀는 동물분류학상 말·얼룩말과 함께 말속(屬)을 형성하고 있다.
- 주 의
 - 특정 동물의 이름은 제외 (예: 곰뱀이, 뱀뱀, 누렁이)
 - 사람, 인간, 호모사피엔스 등은 제외

12.5 AM_AMPHIBIA

- 정 의
 - 양서류
- 예 제
 - 개구리, 두꺼비, 산개구리, 산청개구리, 도롱뇽, 맹꽁이
- 본 문
 - 한국에는 도롱뇽(*Hynobius leechi*)·꼬리치레도롱뇽(*Onychodactylus fischeri*)·네발가락도롱뇽(*Salamandrella keyserlingi*)의 3 속 3 종이 있으며,

12.6 AM_REPTILIA

- 정 의
 - 파충류
- 예 제
 - 도마뱀, 거북, 자라, 구렁이, 백사, 공룡, 악어, 용
- 본 문
 - 우리 나라에는 구렁이에 관한 전설과 전승 속신이 많이 전한다
 - 흔히 민간이나 한방에서는 백사를 희귀한 보약으로 여기고 있으나, 과학적으로는 근거가 없다.
 - 한데 묶어 디노사우르라고 한 것을 동양에서 '공룡'이라고 번역하면서 쓰이기 시작했다.
 - 악어는 담수성이나, 인도악어의 경우에는 바다에도 나온다.

12.7 AM_TYPE

- 정 의
 - 동물 분류 명칭
- 예 제
 - 강장동물, 구두동물, 극피동물, 내항동물, 모악동물, 선형동물, 성구동물, 연체동물, 완보동물, 완족동물, 원생동물, 유선형동물, 유형동물, 윤형동물, 의충동물, 절지동물, 척색동물, 태형동물, 편형동물, 해면동물, 환형동물, 히드로충류, 포유류, 메뚜기과, 어류, 조류, 양서류, 파충류
- 본 문
 - 200 여 종의 조류와 뱀, 개구리, 도마뱀 등 파충류가 서식하고 있다.
 - 실제 크기의 바다 양서류와 공룡 화석, 호주에서 가장 큰 나방이 볼 만하다.
- 주 의
 - 어류/물고기/민물고기도 AM_TYPE 에 해당됨 (AM_FISH, CV_FOOD 로 태깅하지 않음)
 - '새'도 AM_TYPE 에 해당됨
 - '플랑크톤'은 동물 분류 유형에 포함되지 않으므로 AM_TYPE 이 아니라, 생물로 보고 AM_OTHERS 로 태깅함
 - '세균'은 AM_TYPE 으로 태깅한다.

12.8 AM_PART

- 정 의
 - 동물 몸의 한 부분(신체부위) 명칭
 - 일반적으로 눈으로 확인할 수 있는 신체의 겉부분 명칭
- 예 제
 - 팔, 다리, 머리, 장단지, 허벅지
 - 안구, 이빨
- 본 문
 - 갖가지 화려한 색상을 뽐어내는 산호초와, 지느러미와 부리가 용의 모습을 닮은 해룡 (Sea-Dragon), 표범무늬가 독특한 바다뱀 (Sea Snake)은 반드시 보도록 한다.
- 주 의
 - 일반적으로 눈으로 확인할 수 있는 겉부분 명칭만 해당되며, '골반' 같은 경우는 TM_CELL_TISSUE 에 해당함
 - '껍질'은 물질을 나타내므로, AM_PART 에 해당하지 않음
 - "피부"는 외부에선 보이는 부분으로 AM_PART 로도 볼 수 있지만, 생물 조직으로 고려해서 TM_CELL_TISSUE 로 태깅함

12.9 AM_OTHERS

- 정 의
 - 기타 동물 명칭
- 예 제
 - 절지동물(예: 가재), 연체동물(예: 문어, 오징어)
- 본 문
 - 슬리피 라군이라 불리는 산호초들이 파도를 막아주면서 자연적으로 형성된 이스트 아가나 비치에 있는 호텔이다.
 - 사막지대를 달린 후에 스타톤 비치로 내려와서 해변가를 감상하거나, 조개줍기를 하면서 시간을 보낼 수 있다.
- 주 의
 - "세균", "바이러스", "박테리아" 등은 비세포성 생물이므로 AM_OTHERS 에 해당됨
 - "기생충"은 AM_OTHERS 에 해당됨
아가나 비치에 있는 호텔이다.
 - 사막지대를 달린 후에 스타톤 비치로 내려와서 해변가를 감상하거나, 조개줍기를 하면서 시간을 보낼 수 있다.
- 주 의
 - "세균", "바이러스", "박테리아" 등은 비세포성 생물이므로 AM_OTHERS 에 해당됨
 - "기생충"은 AM_OTHERS 에 해당됨

13. PLANT

식물 명칭을 나타낸다. 세분류에 포함되지 않는 식물명은 PT_OTHERS 로 태깅한다.

13.1 PT_FRUIT

- 정 의
 - 과일 이름, 식물의 열매
- 예 제
 - 사과, 배, 자두, 바나나, 복숭아, 오렌지, 포도, 호두
- 본 문
 - 특히 자연지리적으로 과수 재배에 적합하여 복숭아가 많이 생산되며
 - 오렌지의 윗면을 잘라 속을 파낸다.
 - 한국에는 고려시대에 중국에서 들여온 것으로 추측되며 《조선왕조실록》 등에도 포도에 관한 기록이 실려 있다.
 - 호두는 열매가 성숙된 가을에 따서 물에 오랫동안 담가 두거나 한 자리에 쌓아 두어 썩힌 육질의 외과피를 제거하고 햇볕에 말린 뒤 딱딱한 내과피를 깨서 종자를 취한다.
- 주 의
 - '완두콩'은 완두(식물)의 씨앗(열매)이므로 PT_FRUIT 으로 태깅함

13.2 PT_FLOWER

- 정 의
 - 꽃 이름
- 예 제
 - 저먼아이리스, 백합, 국화, 제비꽃, 히아신스, 초롱꽃, 할미꽃
- 본 문
 - 제비꽃은 성실과 겸손을 나타낸다고 하였다.
 - 히아신스는 화단뿐만 아니라 분재나 물재배로도 많이 이용되는데,
 - 화관은 길이 4~5cm 이고 초롱(호롱)같이 생겨 초롱꽃이라고 한다.
 - 흰 털로 덮인 열매의 덩어리가 할머니의 하얀 머리카락같이 보이기 때문에 할미꽃이라는 이름이 붙었다.
- 주 의
 - 정식 꽃 이름 명칭 뒤에 '~꽃(예: 백합꽃)'이 붙어서 나온 경우 통합하여 태깅한다.
 - 제라늄 : 제라늄/PT, 제라늄꽃 : 제라늄 꽃/PT, 튤립 : 튤립/PT, 튤립 꽃 : 튤립 꽃/PT

13.3 PT_TREE

- 정 의
 - 나무 이름
- 예 제
 - 후박나무, 벚나무, 분비나무, 너도밤나무
- 본 문
 - 한국에서는 후박나무 숲을 생활환경으로 하고 있어서 후박나무와 흑비둘기의 분포가 밀접한 관계를 가지고 있으며
 - 부인 일가는 목은 벚나무들이 짙혀 넘어가는 소리를 들으면서
 - 해발고도 1,000~1,200m 위에는 분비나무·가문비나무 등의 아한대성 침엽수림과 활엽수림의 혼합림이 분포한다.

13.4 PT_GRASS

- 정 의
 - 풀 이름
- 예 제
 - 향부자, 하늘타리뿌리, 인삼, 생강, 귀리, 감자
 - 싱아, 갈대, 질경이, 강아지풀, 창포
- 본 문
 - 판이 이 갈대를 꺾어 피리를 만들어 그녀를 그리워하며 불었던 데서 갈대를 음악의 상징으로 여기게 되었다.
 - 수강아지풀(var. gigantea)은 조와 강아지풀의 잡종이다.
 - 뿌리줄기를 창포라 한다.
- 주 의
 - 채소류(예: 양파, 감자, 시금치)도 PT_GRASS 에 해당됨
 - ◆ 토마토도 PT_GRASS 로 태깅한다.
 - 후추는 열매를 양념으로 쓰지만 식물로 고려해서 PT_GRASS 로도 태깅할 수 있음.
 - 음식 재료명으로 쓰였다고 판단되는 경우 AM 또는 PT 에 속하는 개체라 할지라도 CV 로 태깅한다. (무조건적으로 AM, PT 를 우선시하여 태깅하지 않음)

13.5 PT_TYPE

- 정 의
 - 식물 유형 명칭
 - 식물의 세부 종을 구분하는 유형만 대상으로 함
- 예 제
 - ◆ 녹조류, 선태식물, 양치식물, 종자식물
 - ◆ 장미목, 이판화군, 외떡잎식물, 붓꽃과, 백합목, 낙엽교목
 - ◆ 콩과, 참나무과
 - ◆ 여러해살이 수초, 수초, 뿌리채소
- 주 의
 - ◆ 다음은 식물 유형의 대상이 아님
 - ◆ 저지대식물, 고산식물, 꽃, 나무

13.6 PT_PART

- 정 의
 - 눈으로 확인 가능한 식물의 한 부분에 대한 명칭
- 예 제
 - 뿌리, 가지, 가시, 결눈, 꽃, 열매, 잎, 줄기, 꽃가루, 꽃송이, 씨앗
- 본 문
 - 헛뿌리는 고착작용을 하지만 간혹 흡수작용도 한다.
- 주 의
 - PT_PART 는 눈으로 확인 가능한 부분만 대상으로 함. '꽃', '뿌리', '잎', '줄기', '수술', '씨앗' 등 특정 부분의 명칭이 해당됨.
 - '껍질'은 식물에만 존재하는 것인 아닌 일반 물질이므로 PT_PART 에 해당하지 않음
 - '결명자'와 같이 특정 식물의 씨앗이 독자적인 이름을 가진 경우에도 PT_PART 에 해당됨
 - '해바라기씨', '목화씨'는 PT_PART 에 해당됨

13.7 PT_OTHERS

- 정 의
 - 기타 식물 명칭
 - 식물의 세부분류에 해당하지 않는 기타 식물
- 예 제
 - 우산이끼, 버섯, 표고버섯, 백화고
- 본 문
 - 사실 갯지렁이는 모랫속에 들어 있는 미생물과 해초를 먹고 삽니다.
 - 따라서 생선은 적당히 먹되 마늘이나 미역 등 수은해독에 효과 있는 음식을 함께 먹으라는 조언입니다.

14. MATERIAL

MATERIAL 은 물질에 대한 표현을 나타낸다. 물질의 세부 분류에 해당하는 개체명만 대상으로 한다.

14.1 MT_ELEMENT

- 정 의
 - 원소명
- 예 제
 - 이테르븀, 프로메튬, 디스프로슘, 셀렌, 헬륨
- 주의
 - 원소명은 원소기호와 번호를 할당 받은 물질로 한정한다.
 - ◆ 이테르븀 : 주기율표 제 3A 족에 속하는 희토류원소
 - ◆ 크롬 : 주기율표 제 6A 족에 속하는 크롬족원소의 하나
 - 원소기호도 MT_ELEMENT 로 태깅한다.
 - 원소명, 금속물 등 MT 는 해당 원소명 및 금속물 등의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다.

14.2 MT_METAL

- 정 의
 - 금속물
- 예 제
 - 황산철, 알루미늄, 아연, 구리, 수은, 아말감
- 본 문
 - 2 가인 황산철(II)과 3 가인 황산철(III) 및 2 가와 3 가의 철을 모두 함유한 황산철(II) 철(III)이 알려져 있다
 - 알루미늄이라는 이름은 백반(白礬)에서 유래한다
 - 아연과 구리의 합금인 황동(놋쇠)은 옛날부터 알려져 있었으나, 아연 그 자체는 알려지지 않았다. 17 세기경 인도에서 처음으로 금속아연으로 만들어져, 중국을 거쳐 유럽에 전해졌다
 - 구리는 동(銅)이라고도 하며, 은·금과 함께 화폐금속이라고도 한다
 - 6 세기 말경에는 금의 광석에서 금을 추출하는 데에 이미 수은이 이용되었다.
 - 특히 금·은·구리·아연·카드뮴·납 등과 합금을 만들어 아말감이 된다.
- 주 의
 - 복합명사에 포함된 금속물은, 복합어 내부에서도 본래의 금속물적인 의미를 유지하고 있다면 MT_METAL 로 태깅한다.
 - ◆ 예) 마그네슘리본 또는 마그네슘분말로서 카메라 플래시램프·게터·단열재 등에 사용되는 것은 잘 알려진 사실이다.

- 금속물질 중에서 금속 원소는 MT_METAL 과 MT_ELEMENT 가 모두 가능한데, MT_METAL 로 태깅한다 (예: 크롬, 티타늄)
- 준금속 원소는 MT_ELEMENT 로 태깅함 (예: 붕소)

14.3 MT_ROCK

- 정 의
 - 암석 명칭
- 예 제
 - 화강암, 흑운모화강암, 진사, 역암, 사암, 편암, 편마암
- 본 문
 - 장석은 칼륨장석과 사장석으로 이들의 함량 비율에 따라서 화강암을 분류한다. 유색 광물은 무색광물에 비해서 매우 적다. 화강암류는 어느 것이나 미량의 인회석 ·티타나이트 ·갈륨석 ·자철석 등을 포함한다.
 - 황주계층은 점판암, 탄산염암석들로 이루어져 있다. 연화산암군은 흑운모편마산화강암, 우백흑운모화강암 등이 분포한다. 단천암군은 반상흑운모화강암, 중립질흑운모화강암, 세립화강암, 세립질우백화강암, 맥암 들이다.
 - 일반적으로 진사 중에서 좋은 것을 광명사(光明砂)라고 한다.
 - 이러한 쇄설물이 굳어서 암석이 된 것은 각각의 경우에 탁월한 것의 크기에 따라 역암·사암·실트암·점토암이라고 한다.
 - 온도가 비교적 낮은 곳에서 만들어진 편암(片岩)에서는 편리구조(片理構造)를 가지게 되고 고온하에서 만들어진 편마암(片麻岩)에서는 편마상 구조를 이룬다.
- 주 의
 - 암석 명칭의 경우 ~류 등과 결합하여 암석 종류를 나타낼 때 암석만을 태깅 대상으로 한다.
 - ◆ 화강암류 → <MT_ROCK:화강암>류

14.4 MT_CHEMICAL

- 정 의
 - 화학물질 명칭
 - 고체 화학물질, 액체 화학물질, 기체 화학물질
 - 단백질/생화학물질/효소/호르몬
- 주 의
 - 화학물질 중에 원소에 해당하는 경우에는 MT_ELEMENT 와 MT_CHEMICAL 이 모두 가능한데, 원소명에 해당하는 경우에는 무조건 MT_ELEMENT 로 태깅한다.

- 고체 화학물질

■ 예 제

- ◆ 파라핀, 암모니아, 레이온 등
- ◆ 나프탈렌, 수산화나트륨, 염화나트륨, 탄산나트륨
- ◆ 헤모글로빈, 프리온, 홀로수린, 키틴질, 콜레스테롤, 지방, 에스트로겐, 사포닌, 비타민 등

■ 본 문

- ◆ 파라핀은 무색 ·반투명한 고체로 고형(固形)파라핀 ·석랍(石蠟)파라핀이라고 한다
- ◆ 파라핀지(紙) 제조, 전기절연 재료, 크레용의 원료 등으로 사용된다.
- ◆ 유동파라핀은 석유에서 얻는 유탄유의 유분을 잘 정제한 것이며, 무색으로 휘발성이 적으며 거의 냄새가 없는 액체이다
- ◆ 암모니아를 처음 인공적으로 만든 것은 1774 년 J.프리스틀리이다. 대기 중에는 소량 존재하며 천연수 속에도 미량이 존재하는 경우가 있다
- ◆ 종래에는 살충체로서 나프탈렌을 대량으로 생산하였으나,
- ◆ 수산화나트륨의 제조법으로는 염화나트륨 전해법과 탄산나트륨의 가성화법이 있다.
- ◆ 헤모글로빈 한 분자는 4 개의 폴리펩티드 사슬로 되어 있고, 각각의 폴리펩티드 사슬에는 1 개씩의 헴이 함유되어 있다
- ◆ 프리온 단백질의 화학구조에 의해 발생하며
- ◆ 병원체로 보이는 프리온이 뇌 전체에 퍼져 있는 것이 고령의 환자와는 뇌파의 패턴이 다르다는 것이다.
- ◆ 전염인자인 프리온(Prion)에 의해 발병하며, 전염된다
- ◆ 대부분의 해삼류는 몸 안에 홀로수린(holothurin)이라고 하는 사포닌 비슷한 독소를 가지고 있어
- ◆ 몸 표면은 키틴질과 탄산칼슘으로 된 외골격을 가지고 있어 몸을 보호하거나 체내수분의 증발을 막는 데 알맞다.
- ◆ 세포막은 키틴질 또는 셀룰로오스로 이루어진다
- ◆ 코르티코이드는 모두 스테로이드 대사효소의 작용으로 콜레스테롤에서 생합성한다.
- ◆ 콜레스테롤은 인지질과 함께 세포의 막계(膜系)를 구성하는 주요 성분이며
- ◆ 콜레스테롤에서 디히드로에피안드로스테론을 거쳐 합성된다
- ◆ 지방 그대로 직접 장관(腸管)에서 흡수되는 경우도 많다. 흡수된 지방은 일단 간(肝)이나 피하의 결합조직, 장간막(腸間膜), 근육 사이 등에 축적되고, 그 후 필요에 따라 분해되어 에너지원이 된다.
- ◆ 동물성 지방과 식물성 지방으로 분류되는데, 동물성 지방은 동물체 피하의 지방조직이나 장기의 표면에 축적되고, 식물성 지방은 주로 종자에 축적되어 있다.
- ◆ 원인은 염증으로 인해 일어난다고도 하지만 보통 난포호르몬(에스트로겐)의 작용으로 알려졌다.

- ◆ 천연적으로 존재하는 에스트론·에스트라디올·에스테트롤과 이들과 같은 생물활성(生物活性)을 지닌 합성 에스트로겐으로 분류된다.
- ◆ 난소에서 에스트로겐 분비는 하수체전엽(下垂體前葉)에서 분비되는 성선자극(性腺刺激) 호르몬에 의해 지배되는데(하행성 조절)
- ◆ 과거에 열매의 껍질을 비누대용품으로 사용하였으며 사포닌(saponin)이 들어 있다.
- ◆ 적혈구에 대하여 용혈작용을 보이는 것은 적혈구막 속의 콜레스테린이 사포닌과 강하게 결합하여 막구조가 파괴되기 때문이라고 생각되고 있다
- ◆ 예를 들어 비타민 C 는 사람에게는 비타민이지만 토끼나 쥐를 비롯한 대부분의 동물은 몸 속에서 스스로 합성할 수 있으므로 호르몬이다

- 액체 화학물질

■ 예 제

- ◆ 플루오르화수소, 액화석유가스 등
- ◆ 과산화수소, 에탄올, 니트로벤젠, 리놀레산,

■ 본 문

- ◆ LPG 또는 LP 가스로 약칭하며, 또 프로판가스·부탄가스 등으로도 총칭된다
- ◆ 과산화수소는 효소(酵素)인 카탈라아제 또는 과산화효소에 의해 분해된다.
- ◆ 보통 알코올이라고 하면 이 에탄올을 가리킨다.
- ◆ 니트로벤젠은 독성이 강하고 피부에 흡수되기 쉬우므로 취급할 때 조심해야 한다.
- ◆ 공기 중에서 산화되기 쉬운 건성유(乾性油)는 이 리놀레산과 리놀렌산이 주성분이다.

- 기체 화학물질

■ 예 제

- ◆ 황화수소, 이산화탄소, 이산화황 등
- ◆ 오존, 일산화탄소, 일산화질소, 네온

■ 본 문

- ◆ 아황산가스·아황산무수물이라고도 한다
- ◆ 비교적 농도가 높은 오존이 분포하는데, 이것을 오존층이라고 한다.
- ◆ 불완전 연소생성물인 일산화탄소(CO)가 생성된다.
- ◆ 일산화질소는 화학식 NO, 무색 기체로
- ◆ 그리스어 'neos'로부터 네온이라고 명명하였다.

- 주 의

- 화학물질: 본래는 물질 중에서 화학의 대상이 되는 것을 화학물질이라고 하였으나, 현재와 같이 일반용어로서 쓰이는 경우는 화학산업이 만들어내는 물질 또는 천연으로는 없는 물질, 다시 말해서 인공의 물질이라는 뜻으로 쓰이는 수가 많다.
- 즉, 인공으로 만든 화학물질에 대해서 MT_CHEMICAL 대상이 된다.

15. TERM

이 부류는 위에서 정의된 세부 개체명 이외의 개체명 용어를 포함한다. 아래에 정의된 세부 개체명 만을 대상으로 하고, 그 이외의 개체명은 제외한다.

15.1 TM_COLOR

- 정 의
 - 색
- 예 제
 - 흰색, 흰빛, 흑색, 회색, 황, 홍색, 하늘색, 파란빛, 청, 주석백색, 적색, 적갈색, 적, 자주색, 자갈색, 이모색, 은백색, 분홍색, 분백색, 백색, 다갈색, 녹색, 녹갈색, 녹, 노란색, 금색, 검은색, 검은빛, 갈색
- 주 의
 - '검은 자주색' 등과 같이 색깔을 표현하는 형용사와 결합된 경우에도 전체를 TM_COLOR 로 태깅함. 그러나 개체명은 명사나 명사구만을 대상으로 하므로 '검은'만을 태깅하지는 않음.

15.2 TM_DIRECTION

- 정 의
 - 방향을 나타냄
- 예 제
 - 최북단, 중앙부, 중앙, 중부, 서쪽, 서부, 상부, 북쪽, 북서쪽, 북서부, 북부, 북동, 동쪽, 동서, 동북부, 동부, 동남쪽, 동남단, 남쪽, 남서, 남북, 남부, 남동쪽, 남동부, 남동 등
- 주 의
 - '좌우'는 TM_DIRECTION 으로 태깅한다

15.3 TM_CLIMATE

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

- 본 문
 - 동안기후는 서안기후에 비하여 여름에는 고온다습하고, 겨울에는 한랭 건조한 것이 특징이다.
 - W.P.괴펜은 이것을 한대기후(E 기후)와 냉대기후의 경계로 하였다.
- 주 의
 - 특정 기후만을 대상으로 함. "한랭 건조"와 같은 기후 자체는 개체명 대상이 아님

15.4 TM_SHAPE

- 정 의
 - 모양, 형태 명칭
- 예 제
 - 삼각형, 사각형, 팔각형
- 본 문
 - 삼각형의 종류에는 ① 정삼각형: 3 변의 길이가 같은 삼각형, ② 이등변삼각형: 2 변 길이가 같은 삼각형, ③ 직각삼각형: 하나의 내각의 크기가 90°인 삼각형

15.5 TM_CELL_TISSUE

- 정 의
 - 세포/조직, 세포 명칭과 생물의 조직 및 기관에 대한 명칭
 - 눈으로 외부적으로 보이지 않는 동물/신체 및 식물의 조직 및 기관 명칭
- 예 제
 - 근육, 극세포, 간상세포, 딸세포, 신경세포, 표피, 정원세포, 정모세포
 - 항문, 고막, 식도, 골반, 침샘, 고환, 유전자
- 본 문
 - 원추세포와 섞여서 같은 열에 들어서 있으나, 간상세포의 수가 훨씬 더 많다
 - 분열 전의 모세포(母細胞)에 대해서 딸세포라고 하며 낭세포(娘細胞)라고도 한다.
 - 신경세포는 뉴런 전반에 영양을 공급해 주며, 돌기의 수에 따라 무극(apolar) ·단극(unipolar) ·쌍극(bipolar) ·다극(multipolar) 등으로 분류된다.
 - 표피의 주요 기능은 식물체의 보호이다.
 - 정원세포는 유사분열을 거듭해서 일정한 수준까지 증식을 한 다음 성장과 성숙을 거쳐서 정모세포(精母細胞)가 된다.
 - 고막에 도달한 음파는 고막을 진동시키는데 그 진동은 망치뼈·모루뼈·등자뼈와 이소골을 거쳐 내이에 전달된다.
 - 식도의 내강(內腔)은 엄지손가락 굵기 정도이지만, 음식물이 통과할 때는 상당히 확장된다.
 - 골반은 그 속에 성기 ·분비기관 ·소화기관의 하부를 수용하고 보호하는 동시에 양쪽 다리와 연결되어 몸을 떠받치고 있다.

- 파충류 이상에서는 소화액을 분비하는 침샘이 된다.
- 이행부에는 고리 모양의 융기부가 있는데, 이것을 치대(痔帶)·치륜(痔輪)·항문륜(肛門輪)이라고 한다. ('항문륜' 전체 TM 으로 태깅)

- 주 의

- "동맥", "정맥", "모세혈관"은 대상이지만, "혈관"은 대상이 아님
- "피부"는 외부에서 보이는 부분이지만 생물 조직으로 고려해서 TM_CELL_TISSUE 로 태깅함

15.6 TMM_DISEASE

- 정 의

- 증상/증세/질병

- 예 제

- 기침, 폐결핵, 기관지확장증, 레트증후군, 피고름가래, 호흡불규칙, 호르몬 질환
- 고산병, 고소공포증

- 본 문

- 심신이 쇠약하고 폐가 상하여 기침이나 피고름가래가 나오거나 폐결핵·폐농양·기관지확장증이 있는 경우에는
- 초봄의 고산에서 강렬한 자외선의 조사(照射)로 인하여 생기는 급성피부염도 하나의 고산병이다
- 사람을 만나면 얼굴이 붉어져서 만나기를 꺼리는 적면공포증(赤面恐怖症), 끝이 뾰족한 것에 대한 선단공포증(先端恐怖症), 매독에 걸리지 않았나 하는 매독공포증, 고소공포증(高所恐怖症), 이성공포증(異性恐怖症), 밀실공포증(密室恐怖症), 특정동물에 대한 동물공포증 등 그 대상에 따라 헤아릴 수 없을 정도로 많다.

- 주 의

- 질병의 원인인 '바이러스'는 TMM_DISEASE 로 태깅한다.
- '세균'은 질병의 원인으로 볼 수 있지만 AM_TYPE 으로 태깅한다.

15.7 TMM_DRUG

- 정 의

- 약/약품명
- 약품 종류

- 예 제

- 헤파린, 아스피린
- 약품 종류: 소화제, 근이완제, 구충제, 모기향, 백신

- 본 문

- 또한 전신성 혈관내 응고(disseminated intravascular coagulation)와 혈전생성을 억제하며 투여 후 출혈빈도는 헤파린보다 적게 나타난다.
- 고래에서는 인슐린을 비롯하여 고기·심장·간에서 채취하는 엑스류(類) 외에, 건조갑상선·경랍(鯨蠟)·취장건조분말·젤라틴·헤파린 등을 얻을 수 있고
- 근이완제가 이러한 근수축을 방해하는 작용을 한다.
- 근이완제는 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 모기향이다.
- 소화제에는 위장운동을 항진시켜 소화액의 분비를 높이는 약제와 소화액의 분비부족을 보충하는 소화효소제가 있다.
- 작용기전 및 부작용이 아스피린과 비슷하여 흔히 아스피린계 약물이라 칭한다.
- 구충제 복용 후의 완하제(緩下劑)는 충분한 양을 제때에 복용한다.

15.8 TMI_HW

- 정 의

- 하드웨어 용어, 전자기기에 해당하는 제품

- 예 제

- 하드디스크드라이브, 마이크로프로세서, 롬, 디스켓, 자기테이프, 사운드카드, 인공위성
- 컴퓨터, 모니터, 팩스, 카메라, 디카, 디지털카메라, 냉장고, 전자레인지, 핸드폰, 마우스, MP3 플레이어, 전화(기), 노트북, GPS 수신기, TV, dslr 디카, 에어컨, 일반카드 전화기, 텔레비전, 히터기, GPS, ISE 카드, 네비게이션, 방수 카메라, 충전기
- 망원경, 인공위성
- 바코드, QR 코드

- 본 문

- 하드디스크는 자성체를 입힌 원판형 알루미늄 기판을 회전시키면서 자료를 저장하고 읽어 내도록 한 보조기억장치의 하나로서
- 마이크로프로세서는 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(ROM)은 read only memory 의 약칭이다. 컴퓨터의 판독전용 기억장치를 말한다.
- 디스켓은 기록밀도에 따라 2D·2DD·2HD 등으로 분리하고, 크기에 따라 5.25 인치·3.5 인치 등으로 분리하고 있다.
- 자기테이프는 한때 컴퓨터 보조기억장치로 널리 쓰였으나 자료 처리가 순차적으로만 이루어져 처리시간이 느리고, 순차적으로 구성되어 있지 않은 자료를 처리하는 데는 실용적이지 못한 단점이 있다.
- 사운드카드는 4 스피커 지원여부, 3D 사운드 지원여부 리버브나 코러스 같은 이펙트 지원, 풀 듀플렉스 등의 기능에 따라 차이가 난다.

- 주 의

- '아이폰, 갤럭시에스'와 같은 구체적인 하드웨어 상품명은 개념적으로 AF_WARES 에 속하기도 하고 TM_HARDWARE 에 속하기도 함. 따라서 정확한 상품명으로 나온 경우는 AF 로 태깅하고 그 외 일반 명사에 해당하는 제품들은 TM 으로 태깅하는 것으로 구분한다.

- ◆ '아이패드, 아이폰, 갤럭시에스'는 구체적인 상품명이므로 모두 AF 로 태깅함.
- ◆ 스마트폰, TV, 사운드바의 경우 하드웨어의 종류를 나타내는 일반 명사이므로 TM 으로 태깅함.

- TMI_HW 는 '카메라'와 같은 제품 유형이고, TMI_MODEL 은 'D800'과 같은 세부 제품명을 나타냄

- ◆ 예) <초소형 슬라이드 카메라폰:TMI_HW>(<E800:TMI_MODEL>)

- 큐리오시티, 오퍼튜니티(화성 탐사 로봇), 스피릿(화성 탐사 로봇) 등의 특정 로봇 명칭은 TMI_HW 로 태깅함

- '인공위성'은 TMI_HW 에 해당됨

- '바코드', 'QR 코드' 등은 TMI_HW 로 태깅함

- 전문분야(특허)

- 특허 문서에는 특성상 장치명, 모듈명 등이 존재한다. 일반 분야에서는 개체명 대상이 아니지만 특허 분야에서는 ~장치, ~모듈, ~부 등이 중요한 키워드가 되며 개체명 인식의 대상이 됨

- ◆ 예) 이동통신단말기, 무선화상수신부, 회동지지바, 연동지지바, 광섬유격자, 광전송선, 광써큘레이터, 광검출부, 충격측정부 등 => TMI_HW

~장치, ~모듈, ~부 등을 구성하는 부품들도 특정한 HW 인 경우 개체명 대상으로 인식

- ◆ 예) 소자 소켓, 감각 센서, 후각 센서 등 => TMI_HW

- 주의

- 검출기, 단말기와 같이 단어 자체만으로 특정한 개체를 인지하기 어려운 경우는 개체명 대상에서 제외함. 다만, 충전기와 같이 일반적으로 개체가 무엇을 하는 것인지 인지가 가능한 경우는 개체명 대상으로 인식

15.9 TMI_SW

- 정 의

- 소프트웨어 용어

- 예 제

- 트로이목마, 인터넷 웜, 웜바이러스, 님다바이러스 등
- 그래픽유저인터페이스, 리눅스, CAD, C 언어

- 본 문

- 컴퓨터 시스템을 파괴하거나 작업을 지연 또는 방해하는 악성프로그램으로, 웬 외에도 컴퓨터 바이러스, 트로이목마 등이 있다
- 웬의 경우 얼마전까지 PC 상에서는 중요하게 인식하지 않았다
- 그래픽 유저 인터페이스에서는 마우스 등을 이용하여 화면의 메뉴 중에서 하나를 선택하여 작업을 지시한다.
- 리눅스는 소스 코드를 완전 무료로 공개하여 전세계적으로 약 5 백만 명이 넘는 프로그램 개발자 그룹을 형성하게 되었으며, 이들에 의해 단일 운영체제의 독점이 아닌 다수를 위한 공개라는 원칙하에 지속적인 업그레이드가 이루어지고 있다.
- CAD 는 computer aided design 의 약칭으로 컴퓨터를 이용하여 설계하는 것을 뜻하며
- C 언어는 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

- 주 의

- TMI_SW 도 제품/상품이므로 AF_WARES 와 중복되는데, TMI_SW 에 우선적으로 할당
- TMI_SW 는 '운영체제'와 같은 제품유형이고, TMI_MODEL 는 '안드로이드'와 같은 세부 제품명을 나타냄.
- '사이트, 홈페이지, 링크' 등과 같은 일반 인터넷 용어는 태깅하지 않음.

- 전문분야(특허)

- 특허 문서에는 특성상 ~시스템, ~프로그램 등이 존재한다. 일반 분야에서는 개체명 대상이 아니지만 특허 분야에서는 ~시스템, ~프로그램 등이 중요한 키워드가 되며 개체명 인식의 대상이 됨
 - ◆ 예) 데몬, 명함 관리 시스템, 명함 관리 프로그램, => TMI_SW

15.10 TMI_SITE

- 정 의

- url 주소

- 예 제

- <http://www.naver.com>

- 본 문

- 한국야구위원회 인터넷 홈페이지(www.koreabaseball.or.kr) 자유게시판에는 당시 TV 중계를 통해 오심임을 확인한 네티즌들의 분노가 쏟아지고 있다.
- 민예총은 "한국 영화의 부흥을 이끌고 있는 젊은 감독들의 영화 세계를 들여다봄으로써 우리 영화의 현재와 미래를 점검하는 기회를 갖기 위해 강좌를 마련했다"고 밝혔다. ☎02-739-6854, www.myacademy.org

15.11 TMI_EMAIL

- 정 의
 - 이메일 주소
- 예 제
 - aaaa@etri.re.kr

15.12 TMI_MODEL

- 정 의
 - 각종 제품 등의 모델명, 제품의 부품류
- 예 제
 - LM1000-2CX
- 본 문
 - 프로기아가 상급자용 드라이버 TR500 을 출시했다.(제품 번호 모델명이기 때문에 TM)
- 주 의
 - TMI_MODEL 은 TMI_HW/TMI_SW 와 혼동되는 경우가 많은데, 세부 모델명인 경우에만 TMI_MODEL 에 해당됨.
 - TMI_HW 는 '카메라'와 같은 제품 유형이고, TMI_MODEL 은 'D800'과 같은 세부 제품명을 나타냄
 - ◆ 예) <초소형 슬라이드 카메라폰:TMI_HW>(<E800:TMI_MODEL>)

15.13 TMI_SERVICE

- 정 의
 - IT 서비스 용어
- 예 제
 - 와이브로 서비스, DMB 서비스, 위성 DMB
 - 카카오톡, 올레닷컴, FTTH, 엑스박스라이브, 트위터, 오늘의 유머
- 주 의
 - 엑스박스라이브와 같이 서비스만을 대상으로 함 (다른 세부유형과 혼동되는 경우에 다른 세부유형에게 우선권 부여)
 - "팩스서비스"와 같이 일반적인 서비스는 제외함
 - 웹사이트, 홈페이지, 포털사이트 등과 같은 일반적인 인터넷 용어는 태깅하지 않는다.
- 전문분야(특허)
 - 아날로그 방송, 디지털 방송은 특허분야에서는 특정한 서비스를 의미하는 확률이 높기 때문에 TMI_SERVICE 로 태깅함

15.14 TMI_PROJECT

- 정 의
 - IT 프로젝트 용어
 - 국가 프로젝트, 회사 프로젝트에 모두 사용
- 예 제
 - BrainKorea21, 4 대강사업, 그린 IT 사업, 무료치과진료사업 등
- 주 의
 - IT 계열에 포함되지만, IT 외의 국가/회사 프로젝트에 모두 사용
 - 머큐리 계획과 같이 우주관련 계획은 IT 프로젝트라 보고 TMI_PROJECT 로 태깅

15.15 TMIG_GENRE

- 정 의
 - 게임 장르
- 예 제
 - RPG, 실시간 전략 게임, 1 인칭 슈팅 게임
 - 콘솔 게임, 비디오 게임, PC 게임, 모바일 게임, 온라인 게임
- 본 문
 - 이미 일본과 세계 각지에서 인기를 끌고 있는 'Wii'는 이미 국내에서 140 만대 이상이 팔린 '닌텐도 DS Lite'의 인기를 힘입어 국내 콘솔 게임 시장의 변화를 예고하고 있다.
 - 체감형 스포츠 게임 'Wii 스포츠'의 CF 로 'Wii' 캠페인 모델 활동을 시작한 원빈은 영화 '태극기 휘날리며' '우리형', 드라마 '가을동화' 등으로 잘 알려진 배우로 잘생긴 외모와 뛰어난 연기력으로 국내는 물론 아시아 지역에서도 큰 인기를 끌고 있다.

15.16 TM_SPORTS

- 정 의
 - 스포츠/레저 용어 (기술, 규칙 이름 등)
- 예 제
 - 밀어내기 득점, 솔로 홈런, 바스켓카운트, 리바운드, 발리슛, 삼진, 볼넷, 사구, 사사구, 1 점 홈런, 투런포, 만루홈런, 삼구 삼진, 루킹 삼진
- 본 문
 - 1~1.5m 내외의 퍼팅을 수도 없이 놓쳤다.
 - 특히 세트피스 상황에서 득점력이 좋지 않았던 FC 서울로서는 원발 킥이 뛰어난 무삼파의 가세로 더욱 다양한 공격 루트를 보유할 수 있게 됐다.

- 주 의

- 방향성 정보+땅볼/안타/파울/홈런 등은 묶어서 TM_SPORTS 로 할당함
 - ◆ 예) 1 루 강습 안타, 3 루 땅볼, 좌전 안타, 3 루 펜스쪽 파울, 중월홈런
- 볼넷, 삼진과 같은 스포츠 규칙 관련 용어나 수량의 의미가 없는 1, 2 루와 1, 2 레인은 QT 가 아닌 TM_SPORTS 로 태깅한다.
- 1 볼넷 : QT
- 6 이닝은 QT 로 태깅한다.
- 3 회 말, 9 회 초는 문맥에 따라 **TI** 로 태깅한다.

다. 태깅 단위와 태깅 범주 문제

여기에서는 개체명을 태깅할 때 주의해야 할 두 가지 문제, 즉 개체명 태깅 단위와 둘 이상의 의미 범주에 속하는 개체명을 태깅하는 원칙에 대해 상론한다.

1. 태깅 단위

1.1 구 단위의 태깅

개체명은 명사 또는 명사구를 대상으로 하므로 명사구 전체를 한번에 태깅할 것인지, 단어 단위로 나누어서 태깅할 것인지 문제가 되는 경우가 있다. 이 경우, '최소 단위 태깅 원칙'과 '최장 단위 태깅 원칙'에 준하여 태깅한다.

1.1.1 최소 단위 태깅 원칙

- 각 단어 단위로 나누어서 태깅해도 전체 명사구의 의미가 유지되는 경우 '최소 단위 태깅 원칙'에 따라 각 단어를 따로 태깅한다. 지역명(LC)의 연속 혹은 지역명과 그곳에 위치한 건축물(AF)이나 그 지역을 본거지로 하는 조직(OG)은 각기 따로 태깅해도 전체 구의 의미를 유지하므로 하나로 태깅하지 않는다. 날짜(DT)나 횟수(QT)가 어떤 행사(EV)나 상(CV)과 함께 쓰일 때에도, 각기 태깅해도 전체 구의 의미를 유지하므로 구분하여 태깅한다.
- 예제
- LC + LC
 - ◆ 대전 유성 신성동
 - <대전/LC> <유성구/LC> <신성동/LC>
 - ✕ <대전 유성구 신성동/LC>
- LC + AF
 - ◆ 잠실 롯데월드타워
 - <잠실/LC> <롯데월드타워/AF>
 - ✕ <잠실 롯데월드타워/AF>
- LC + OG
 - ◆ 미국 컬럼비아 대학
 - <미국/LC> <컬럼비아 대학/OG>
 - ✕ <미국 컬럼비아 대학/OG>
 - ◆ 대한제국 법부 전라북도재판소
 - <대한제국/LC> <법부/OG> <전라북도재판소/OG>
 - ✕ <대한제국 법부 전라북도재판소/OG>
- OG + AF
 - ◆ 현대오일뱅크 대산공장
 - <현대오일뱅크/OG> <대산공장/AF>
 - ✕ <현대오일뱅크 대산공장>

- DT, QT + CV, EV
 - 2000 년 제 34 회 납세자의 날 대통령상
 - <2000 년/DT> <제 34 회/QT> <납세자의 날/DT> <대통령상/CV>
 - ✕ <2000 년/DT> <제 34 회 납세자의 날/DT> <대통령상/CV>
 - ✕ <2000 년/DT> <제 34 회 납세자의 날 대통령상/CV>
 - 제 10 회 경제정의기업상
 - <제 10 회/QT> <경제정의기업상/CV>
 - ✕ <제 10 회 경제정의기업상/CV>
 - 1988 년 서울 올림픽 경기
 - <1988 년/DT> <서울 올림픽 경기/EV>
 - ✕ <1988 년 서울 올림픽 경기/EV>

1.1.2 최장 단위 태깅 원칙

- 단어 단위로 나누어 태깅하면 전체 구가 갖는 의미를 나타내지 못하는 경우에는 최장 단위를 태깅 대상으로 한다.
- 예를 들어, '신고전학과 성장론'은 그 전체가 이론 명칭이다. 이 명사구 내부에는 '신고전학과'라는 FD 가 포함되어 있으나 '성장론'만으로는 특정한 이론을 지시할 수 없게 되므로 이러한 경우에는 '신고전학과 성장론' 전체를 TR 로 태깅한다.
- 예제
 - 88 서울 올림픽 경기
 - 1988 년이 줄여서 '88'이라고 표현된 경우 '88'만으로는 1988 년임을 보여줄 수 없다. 따라서 개최 연도가 축약되어 표기된 경우는 후행하는 EV 와 함께 태깅한다.
 - <88 서울 올림픽 경기/EV>
 - ✕ <88/DT> <서울 올림픽 경기/EV>
 - 한국해양대학
 - '한국해양대학'의 경우 대학 이름에 '한국'이라는 지명이 포함된 것이므로 함께 태깅하지 않으면 본래 단체명이 훼손된다.
 - <한국 해양 대학/OG>
 - ✕ <한국/LC> <해양 대학/OG>
 - 스페인 대사관
 - '스페인 대사관'은 대사관이 소재하고 있는 위치를 의미하는 것이아니므로 통합하여 태깅한다.
 - <스페인 대사관/OG>
 - ✕ <스페인/LC> <대사관/OG>
 - ◆ 제 24 회 서울 올림픽 경기
 - 스포츠 행사는 보통 개최 횟수와 개최지가 함께 출현하는데, 개최 회수(QT)는 구분하여 태깅하나 행사가 열리는 지역(LC)은 통합하여 태깅한다.
 - <제 24 회/QT> <서울 올림픽 경기/EV>
 - ✕ <제 24 회/QT> <서울/LC> <올림픽 경기/EV>
 - ✕ <제 24 회 서울 올림픽 경기/EV>

1.2. 복합어 내부 태깅

- 둘 이상의 단어가 만나 합성어를 이루거나 단어와 접사가 만나 파생어를 이룬 경우 그 전체가 하나의 개념을 지시한다고 볼 수 있으므로, 복합어 전체가 개체명일 경우에만 태깅하고 복합어 내부 요소는 개체명이더라도 별도로 태깅하지 않는다.
- 예제
 - ◆ 손바닥: 복합어 전체가 개체명이므로 전체를 AM 으로 태깅함.
 - ◆ 손짓: 복합어 전체는 개체명이 아니므로 '손짓'은 태깅하지 않음. 내부 요소 '손'도 AM 으로 태깅하지 않음.
 - ◆ 야구장: 복합어 전체는 특정한 인공물을 지시하지 않아 AF 가 될 수 없어 태깅하지 않고 내부 요소인 '야구'도 CV 로 태깅하지 않음.
 - ◆ 한국적: '한국적 문화'와 같은 표현에서 '한국적'은 개체명이 아니므로 태깅하지 않고 내부 요소 '한국' 역시 LC 로 태깅하지 않음.
 - ◆ 국어국문학과: '국어국문학과'는 어떤 대학교의 하위 부서이므로 OG 로 태깅하지 않고 내부 요소인 '국어국문학' 역시 FD 로 태깅하지 않음.
- 다만, 다음의 경우는 예외적으로 복합어 내부 요소를 태깅하므로 주의를 요한다.
- 원소명, 금속물 등 MT 는 해당 원소명 및 금속물 등의 의미를 유지하고 있을 때 복합어 내부 요소임에도 태깅할 수 있다.
 - ◆ (예) 마그네슘리본: <마그네슘/MT>리본
- 다음의 접미사가 결합하여 만들어진 복합어는 예외적으로 내부 요소를 태깅한다.
- 첫째, 접미사 '-들'은 선행 요소의 의미에 영향을 주지 않으므로 '-들'이 결합한 단어는 '-들'을 제외하고 태깅한다.
 - ◆ (예) 학생들: <학생/CV>들
- 둘째, 접미사 '-님'이 결합한 단어는 전체가 <우리말샘>에 등재된 경우(즉, 관습적으로 '-님'을 붙여서 직위·직책명으로 사용하는 경우) 전체를 태깅하나 그렇지 않은 경우는 '-님'을 제외하고 태깅한다.
 - ◆ (예) 선생님: <선생님/CV>, 회장님: <회장/CV>님
- 셋째, 낱파(DT), 시간(TM), 수량(QT) 표현은 접미사('-'당, -씩, -쯤)가 결합하여 의미를 구체화시키는 경우 접미사까지 통합하여 함께 태깅한다.
 - ◆ (예) 한 명당: <한 명당/QT>, 1 미터당: <1 미터당/QT>
 - ◆ (예) 삼 년쯤: <삼 년쯤/DT>
 - ◆ (예) 세 시간씩: <세 시간씩/TM>

2. 유일 태깅 원칙

- 개체명 중에는 둘 이상의 의미 영역에 동시에 속하여서 중복 태깅이 가능한 경우가 있다. 본 지침은 그 경우에도 단일한 태그만을 부여하는 태깅 원칙을 갖는다. 따라서 중복 태깅이 가능한 개체명 단어가 출현했을 경우에는 우선순위를 정하여 하나의 태그만을 할당해야 한다. 이를 '유일 태깅 원칙'이라 한다.
- 유일 태깅 원칙을 적용해야 하는 대상에는 두 가지 유형이 있다. 첫 번째는 문맥과 관계없이 해당 개체가 본유적으로 두 가지 의미 범주에 속하는 경우이다. 이 경우는 지침에서 정해 둔 우선순위에 따라 일관적으로 태그를 부여한다. 두 번째는 문맥에 따라 의미 범주가 달라질 수 있는 부류이다. 이 경우는 문맥을 우선적으로 적용하여 태깅하되, 문맥으로 판단이 어려운 경우에만 우선순위를 적용하여 태깅한다.

2.1 우선순위 적용 유형(유일 태깅)

- 아래의 유형은 문맥을 고려하지 않고 일관적으로 우선순위에 따라 태깅해야 한다. 예를 들어 '경포대'의 경우 관광지로서 LC_TOUR 에 속하며 건축물로서 AF_BUILDING 에 속하기도 하고 문화재로서 AF_CULTURAL_ASSET 에 속하기도 하는데, 이는 문맥에 따라 달라지는 것이 아니라 '경포대'가 지닌 본유적인 속성이다. '경포대'의 경우 'AF_BUILDING < AF_CULTURAL_ASSET < LC_TOUR'라는 우선순위에 따라 LC_TOUR 로 태깅한다.

중복유형		우선순위	예제
태그 1	태그 2		
AF_CULTURAL_ASSET	AF_BUILDING	AF_CULTURAL_ASSET	경복궁, 보현사 등
AF_CULTURAL_ASSET	AF_WORKS	AF_WORKS	고려사, 신증동국여지승람, 동의보감, 인왕제색도
AF_CULTURAL_ASSET	LC_TOUR	LC_TOUR	천연기념물(예: "단양영천리 측백나무숲", 담양관방제림")
LC_TOUR	LC_PLACE	LC_TOUR	경포대 해수욕장
LC_TOUR	AF_BUILDING	LC_TOUR	경포대
MT_ELEMENT	MT_METAL	MT_METAL	크롬, 칼륨, 티타늄
MT_ELEMENT	MT_CHEMICAL	MT_ELEMENT	산소, 질소, 규소, 탄소
AF_WARES	TMI_HW	TMI_HW	카메라, 핸드폰, 인공위성
AF_WARES	TMI_MODEL	TMI_MODEL	D800
AF_WARES	CV_FOOD	AF_WARES	백종원 도시락
AF_WARES	CV_DRINK	CV_DRINK	연꽃차

AF_WARES	CV_CLOTHING	CV_CLOTHING	오리털재킷
----------	-------------	-------------	-------

2.2. 문맥 우선 적용 유형

- ☞ 무조건적으로 우선순위를 적용해야 하는 유형과는 달리 문맥에 의해 중의성이 해소되는 유형도 있다. 이 경우는 문맥에 의존하여 개체명 부류를 판단하고 문맥으로 판단이 불가능할 경우에만 다음과 같은 우선순위를 적용하여 태깅한다.

AM_FISH	CV_FOOD	AM_FISH	명태, 고등어
PT_GRASS	CV_FOOD	PT_GRASS	시금치, 오이, 철

- 위의 원칙에서 음식과 동식물에 관해서는 무조건적으로 우선순위를 적용하는 것이 아닌 문맥을 고려하여 따져봐도 애매한 경우에만 우선순위를 적용하여 태깅한다.
 - ◆ '오늘은 시금치를 다듬어서 파스타를 만들어 볼 건데요.'와 같은 경우 '시금치'는 음식 재료로 나왔기 때문에 '시금치/CV'로 태깅한다.
 - ◆ 하지만 '시금치 금값되나'와 같이 재료인지, 식물인지 판단이 애매할 경우 '시금치/PT'로 태깅한다.
- LC, AF, OG 중복 태깅이 가능한 경우 역시 문맥에 따라 판단을 한다. 문맥에 따라 OG가 아닌 AF로 판단이 되는 경우 AF로 태깅을 하며, AF가 아닌 OG로 판단이 되는 경우 OG로 태깅한다. (LC의 경우도 동일하게 적용함)
- '국가'나 '도시', '행정구역명'을 LC 또는 OG로 판단하는 기준은 문맥으로, 해당 문장의 서술어를 중심으로 한 문장 의미에 따라 결정한다.
 - ◆ 예) 북한은 한반도 비핵화를 위해 노력하고 있다며 → 북한/OG
- 단, '명사(국가/도시/행정구역명)+명사' 구성에서 앞부분에 해당하는 '명사(국가 등)'는 LC로 일괄 태깅한다.
 - 예) 서울 시장 → 서울/LC, 시장/CV
- 문맥에 따라 판단이 어렵거나 애매한 경우 해당 개체명이 속한 원래의 태그를 최우선으로 적용한다.
 - 예) A 씨는 실종 이틀 만인 25 일 오전 9 시 50 분께 우암산 보현사 인근에서 중앙구조본부 인명 구조견에 의해 발견됐다.
→ LC, AF 판단이 어려움, 사찰의 경우 원래 AF에 속해있기 때문에 AF로 태깅
- 그러나 '개체명이 속한 원래의 태그'라는 기준은 현 지침에서 명시되어 있는 범위에 한정되므로, 지침에 명시되지 않은 부류가 나올 경우 특정 건물을 의미하는 사례는 AF로 '단지' 등 더 넓은 공간을 의미하는 사례는 LC로 태깅한다.

라. 기타

1. 개체명 성능 평가

개체명 성능평가는 전통적으로 F-score 를 사용한다[Sang 과 Meulder, 2003].

평가 대상은 앞장에서 설명한 태그별 태깅 기준에 따라 구축된 두개의 데이터 (annotation result, answer)이다. Answer 는 평가의 기준이 되는 데이터를 의미하고, annotation result 는 기준 데이터 answer 와 성능을 비교할 데이터이다.

F1 measure 은 F-score 에서 두가지의 평가 수식(Precision, Recall)의 조화 평균을 구하는데 이 비율을 같게 계산하는 방법이다. 개체명 성능 평가에 사용할 F1 measure 의 수식은 다음과 같다(수식에서 annotation result 를 result 로 표시).

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

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

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Precision 은 annotation result 데이터 개체명 중 기준 answer 와 일치하는 비율을 말하고, 반대로 Recall 은 answer 의 개체명 중 annotation result 와 일치하는 비율을 뜻한다. 이들의 조화평균이 F1 이다.

일치는 개체명이 문장(또는 문서) 내 시작과 끝 위치를 비교한다. 예를 들어 "2006 년 보다 200 만원 하락"라는 문장에서 annotation result 와 answer 의 결과 json 이 다음과 같은 경우 일치는 1 개 이다. Json 의 begin, end 를 모두 일치하는 것이다.

Annotation result
<pre>"NE": [{"id": 1, "form": "2006 년", "label": "DT", "begin": 0, "end": 5}, {"id": 1, "form": "200", "label": "QT", "begin": 9, "end": 11}]</pre>

Answer
<pre>"NE": [{"id": 1, "form": "2006 년", "label": "DT", "begin": 0, "end": 5}, {"id": 1, "form": "200 만원", "label": "QT", "begin": 9, "end": 14}]</pre>

2. 기타 논의 사항

2.1. 원어 정보의 경우

- 원어 정보가 개체명 사이에 괄호로 일부만 제시된 경우는 괄호를 제외하고 앞뒤를 합쳐야만 개체명이 되기 때문에 전체를 묶어 괄호까지 포함하여 태깅한다.
- 예 제
 - 퓨 (Pew)공익재단의 → 퓨(Pew)공익재단/OG
 - 테링 (鐵嶺)시 → 테링(鐵嶺)시/LC
- '이코노미스트지(誌)'와 '윤덕여호(號)'처럼 원어 정보가 개체명 끝부분에 괄호로 일부만 나왔을 때는 괄호 정보를 제외해도 하나의 개체명으로 인식이 가능하기 때문에 괄호정보를 태그하지 않는다.
- 예 제
 - 이코노미스트지(誌)는 → 이코노미스트지/OG
 - 윤덕여호(號) → 윤덕여호/OG
- '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/OG+문예위/OG
 - 미국 드라마(미드) → 미국 드라마/FD+미드/FD
- 주의
 - 개체명과 함께 괄호 정보가 개체명 원어 정보 전체를 나타낼 경우는 괄호 정보를 태깅하지 않는 것으로 수정함.
 - 후진타오(胡錦濤) → 후진타오/PS +胡錦濤/PS ✕
 - 후진타오(胡錦濤) → 후진타오/PS ✓
 - 수월관음도(水月觀音圖) → 수월관음도/AF

2.2. 원어 정보가 아닌 부가 정보가 괄호로 제시된 경우

- 괄호 안의 부가 정보가 개체명일 경우에는 괄호 안의 정보도 태깅한다.
- '이름(국가, 순위, 나이)'로 나온 경우 괄호안의 정보를 각각 '국가/LC, 순위/QT, 나이/QT'로 태깅한다.
- 예제
 - 샤라포바(러시아, 2 위, 28 세) → 샤라포바/PS+러시아/LC+2 위/QT+28 세/QT
 - 이번 불화(세로 123.5cm×가로 60.5cm) → 불화/FD+세로 123.5cm×가로 60.5cm/QT
- 주의
 - 단, 괄호 안의 부가 정보가 개체명 태깅 범위에 속하지 않는 경우에는 괄호 정보 앞의 개체명만 태깅한다.
 - 가운데점이나 여타의 기호가 개체명 안에 포함되어 있는 경우 함께 태깅한다.
 - ◆ 5·8 시 → 5·8 시 /TI 취득·등록세 → 취득·등록세/CV
 - 괄호 정보 속 기호 'x'도 개체명 내부에 있는 경우 함께 태깅한다.

2.3. 축소 지칭의 경우

- 텍스트 내 다른 문장, 문맥을 고려하지 않고도 단독으로도 줄인 말로 쓰이는 경우(보편적으로 줄인 말로 자주 사용되는 경우) 복원된 형태에 준하여 태깅한다.
 - '수원 삼성' (중략) 수원이 3 대 0 으로 울산을 이겼다.
→ '수원 삼성'의 경우, 사람들은 선행 문장이 없어도 "어제 수원이 3 대 0 으로 이겼어"라고 말할 수 있으며 실제로도 지역명만 팀 이름으로 자주 불리는 경우를 신문, 뉴스 등에서 쉽게 찾아 볼 수 있음.
→ 따라서 수원/LC 가 아닌 수원/OG 로 태깅
 - 과천시립미술관 ... (중략) 미술관
→ 문맥 전체를 보고 판단해야 미술관이 '과천시립미술관'이라는 것을 알 수 있지 문장 안에서 단독으로 쓰였을 경우 특정성이 없기 때문에 태깅할 수 없음.
→ 미술관 태깅하지 않음. (과천시립미술관은 태깅함)
 - 부산국제영화제 (중략) 부산
→ 앞 문맥 또는 문장 안에서 '부산'이 부산국제영화제를 지칭한다고 보일지라도 '부산'은 보편적으로 영화제라고 인식하지 않음. 홀로 쓰였을 때 '부산국제영화제'라고 쓰인다고 보기 어렵기 때문에 부산/EV 로 태깅하지 않음.
- 즉, 어떠한 단어나 구에서 선행 혹은 후행 요소가 절단된 상태의 단어가 텍스트 내 다른 문장이나 문맥 없이도 본래 어형의 의미 범주에 속하는 경우, 본래 어형의 의미 범주에 준하여 태깅한다. (예: 수원 삼성/OG → 수원/OG)
- 텍스트 내 다른 문장이나 해당 문장의 다른 단어를 고려해야만 어형이 복원되는 경우는 홀로는 그 의미를 지니지 못한다고 보고 절단된 단어 자체의 의미 범주에 준하여 태깅한다. (예: '부산국제영화제/EV → 부산'에서 '부산'은 문맥 없이 단독으로는 EV 의 의미를 지니지 못하므로 '부산'이 갖는 본래 범주인 LC 혹은 OG 로 태깅하게 됨. '칸영화제/EV → 칸'에서도 '칸'을 EV 로 태깅하지 않음.)
- 예 제
 - FC 서울이 이번 경기로 ...(중략), 서울은 ...(중략). → FC 서울 /OG, 서울/OG

2.4. 일종의 '비유적 표현'의 경우

- 비유적 표현이라 할지라도 비유인지 아닌지를 고려하지 않고 해당 개체명 분류에 알맞게 태깅한다. (보이는 그대로 태깅)
- 예 제
 - 민주당의 입장에서는 '않던 이'가 빠진 격이다. → 이/AM
 - 그는 결혼과 취직이라는 '두 마리 토끼'를 동시에 잡았다. → 두 마리/QT + 토끼/ AM

2.5. [날짜/시간/수량 N + 보조사/접사/단위성 의존명사 지칭 변동 사항

- 날짜, 수량, 시간과 관련하여 보조사, 접사, 단위성 의존명사 등과 함께 나와 의미를 구체화시키는 경우 통합하여 함께 태깅하는 것으로 수정한다.
- 따라서 '만, 쯤, 씩' 등도 포함하여 태깅한다.
-
- 예 제
 - 10 일 가량/DT
 - 10 억원대/QT
 - 3 일째/DT
- 주의
 - 날짜/수량/시간 N + 일반 명사 : 정도, 가량, 이상, 이하, 이내 등과 같은 경우 포함하여 태깅한다. 예) 30 일 정도/DT, 2km 이내/QT
 - '연속'은 태깅하지 않는다.

2.6. AF 경계가 불명확한 경우

- '돔구장, 야구장' 등은 특정한 건축물을 지시하지 못하므로 태깅하지 않는다.
- 아울러 '광주 돔구장' 및 '서울 올림픽공원 체조경기장'의 경우 정식 명칭이 아니므로 전체를 한번에 태깅할 수 없다. '광주/LC', '서울/LC, 올림픽공원/LC'만 태깅하며 '돔구장'과 '체조경기장'은 태깅하지 않는다.
- 'LC+AF'와 'OG+AF'는 LC 및 OG 와 AF 를 따로 태깅한다. 다만 AF 는 단독으로 특정한 건축물을 지시할 수 있을 때만 태깅한다.
- 예 제
 - 현대중공업 조선소: 현대중공업/OG
 - 현대오일뱅크 대산공장: 현대오일뱅크/OG, 대산공장/AF
 - 광주 돔구장: 광주/LC
 - 서울 올림픽공원 체조경기장: 서울/LC, 올림픽공원/LC
 - 스타벅스 여의도점: 스타벅스 여의도점/OG

3. 구어 말뭉치 태깅

3.1. 오타의 경우

- 원시어절에서 명백한 오타(전사상 오류, 발음 오류 등)로 보이는 경우에도 개체명 분류에 속하면 원어절 그대로 해당 개체명으로 태깅한다.
- 원문을 수정해서 태깅하지 않는다. (후처리로도 원문을 수정하지 않음.)
- 예 제
 - 어~ 자기가 전혀 모르는 언어 지금 <trunc>까</trunc> 지금 저기 뭐 <trunc>오</trunc> <trunc>오</trunc> 오스트릴라에 → 오스트릴라/LC 로 태깅
 - 사라코지 대통령이 → 사라코지/PS 대통령/CV

3.2. 줄임 표현의 경우

- 줄임 표현의 경우 개체명이 포함되어 있는 경우 형태소 제거가 불가하다면 일단 전체를 개체명으로 태깅하고 메모를 달아준다. (이번 말뭉치 분석 작업에서만 이렇게 작업함. 후처리로 제거 필요)
- 예 제
 - 삼십프로가? → '삼십프로/QT'지만 '삼십프로/QT'으로 태깅
 - 러시아안데요. → '러시아/LC'지만 '러시아/LC'으로 태깅
 - 하날니다. → '하날/QT'지만 일단 '하날/QT'으로 태깅
 -

3.3. 원어절 중간에 '~' 와 같은 기호가 삽입되어 있는 경우

- '이천 이~ 십년'과 같이 중간에 '~'가 삽입되어 있는 경우에는 일단 '~'를 포함하여 태깅한다.
- '이~'는 형태 분석 시에는 담화 표지로 인식된다. 따라서 일단 포함하여 태깅하되, 의미가 없는 것으로 판단한다.
- 예 제
 - 이천 이~ 십년 오월과 칠월 : 이천 이~ 십년/DT, 오월/DT, 칠월/DT

Analysis of named entity and corpus construction

This project is to build a large-scale, high-quality corpus on the named entity that can be used as a basis for the development of Korean language processing technology for the Fourth Industrial Revolution. We aimed to build a large-scale, high-quality corpus through rigorous verification and refinement of the program and experts suitable for constructing an named entity corpus. In addition, we have been familiar with the ETRI guidelines used to build a corpus, and have also attempted to supplement the guidelines with the National Institute of Korean Language by reflecting the revisions and supplements of the guidelines.

The project takes about four months, and the target is 3 million words of raw corpus delivered from the National Institute of Korean Language. Data.ly and Changwon National University, which was a joint project partner, received 2 million words of written raw corpus and 1 million words of spoken raw corpus from the National Institute of Korean Language. We aimed at an error rate of less than 5%.

In order to achieve successful results efficiently, the project proceeded as follows.

First, we utilized various programs owned by Data.ly. We performed the work through an automated system, such as a labeling and verification tool based on consistency checks, an error determination method using patented technology, Rough set, and an automatic error correction method based on RDR. After automatically attaching the 15 named entity categories set out in the ETRI guidelines using an automated system, we constructed, refined the corpus, and detected errors. The operation process of the automation system is carried out collectively under the control of the expert from error detection to correction, thus obtaining both expertise and accuracy. We reviewed and revised the results produced by this work by the experts once again. In addition, we reused all the results produced by the workers for improving the performance of the automation system. Using the error-correction pairs that occurred during the operation, we could automatically obtain the error correction rules and apply them to the whole corpus in the system to make batch corrections and finally reduce the errors.

Second, we hired workers to build and modify the corpus, to review the same data three times. All workers were trained by specialists before beginning this work and had to go through the guidelines through annotating the sample corpus. We use Data.ly's program "Kane" as a tool for building named entity corpus, increasing work efficiency and reducing simple mistakes. The workers reviewed and corrected the preprocessed corpus through an automated system. In addition, experts responded to and shared the questions of the workers, and repeatedly retrained the workers on the guidelines. At the same time, a quality assurance team centered on Korean language specialists was set up to review and revise the results of primary workers. Through this, the error rate was reduced and the corpus was completed. The opinions of experts generated during the review were summarized so that they could be reflected in the guidelines.

Finally, we forwarded all of the team's outputs, the named entity corpus, and any query or guideline modifications that occurred during the work to the National Institute of Korean Language for review. After receiving the errors and fixes for the named entity corpus from the National Institute of Korean Language, we proceeded to correct them. For issues that need to be discussed, the opinions are adjusted and then reflected in the guidelines to improve the completeness of the guidelines.

The corpus completed through this project will be used as a common material for everyone to use freely, thereby becoming a foundation for developing language processing technology using artificial intelligence. It also has the significance of establishing basic data for the study of Korean language and culture. We expect that the discussions arising from the course of the project will help in the development of Korean language research policies.

The corpus construction task is attracting a lot of attention, and its importance will increase in the future. Data.ly and Changwon National University participated in the project "Building a named entity corpus" ordered by the National Institute of Korean Language to assist in the development of Korean language research and related industries.

Keywords: Automation system, guideline, corpus, validation, instructions

Project Director: Cha Jeong Won (Data.ly)

사업 책임자 차정원 (데이터리 대표)

사업 참여자 정경재 (창원대학교 국어국문학과 조교수)
 오진영 (데이터리 이사)
 박은실 (창원대학교 산업기술연구원 전임연구원)
 배진영 (데이터리 전임연구원)
 최윤진 (데이터리 전임연구원)

담당 연구원 이승재(국립국어원 언어정보과장)
 이수미(국립국어원 언어정보과 학예연구사)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9718, 전송 02-2669-9727

인쇄일: 2019년 11월 22일

발행일: 2019년 11월 22일

인 쇄: 서경기획

※ 이 책은 국립국어원의 용역비로 수행한 ‘개체명 분석 말뭉치 구축’ 사업의
결과물을 발간한 것입니다.