



한국해양대학교 - 컴퓨터공학과
자연언어처리 연구실

부산광역시 영도구 태종로727
051-410-4896

한국어 개체명 정의
(Definition of Korean Named-Entity Task)

(ver 1.0)

TR No: KMOU-NLP-2018-001

한국해양대학교 자연언어처리연구실
2018년 10월

한국어 개체명 정의*

(Definition of Korean Named-Entity Task)

한국해양대학교 자연언어처리연구실

천민아, 김재훈

부산광역시 영도구 태종로727 한국해양대학교 컴퓨터공학과

✉ kmounlp@naver.com

☎ 051)410-4896

한국어에서 개체의 범주는 크게 **개체이름**, **시간표현**, **수량표현**으로 분류할 수 있다. **개체이름**은 인명(PER), 지명(LOC), 기관명(ORG), 기타(POH)로 나누었고, **시간표현**은 날짜(DAT), 시간(TIM), 기간(DUR)로 나누었으며, **수량표현**은 통화(MNY), 비율(PNT), 기타 수량표현(NOH)으로 나누었다.

* 이 보고서는 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(R7119-16-1001, 지식증강형 실시간 동시통역 원천기술 개발)과 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017M3C4A7068187)

1. 한국어 개체명 인식 연구범위

가. 개체명 인식(Named-Entity Recognition)

개체명 인식은 크게 개체이름 인식, 시간성 표현 인식, 수량표현 인식으로 구성된다. **개체이름**은 고유명사에 해당하는 것으로 실세계에서 유일한 개체를 표현하는 이름 혹은 이름을 대신할 수 있는 표현이다. 인명, 지명, 기관명 등이 개체이름에 해당한다. **시간성 표현**은 날짜, 시간, 기간 등을 표현하는 구절이다. **수량 표현**은 단순한 숫자, 금액과 비율 등 수량을 표현할 수 있는 구절이다.

나. 한국어 개체명 정의의 원칙

한국어 개체명 정의 및 개체명 인식을 위한 말뭉치 구축을 위한 다섯 가지 원칙을 소개한다.

(1) 개체명은 단어를 중심으로 한다.

어떤 개체명이 형태소로 분리되는 경우에는 개체명으로 간주하지 않는다. 예를 들어, ‘한국인’의 경우 ‘한국+인’으로 분리될 수 있다. ‘한국’은 위치(LOC)를 나타내는 개체명 혹은 국가(ORG)를 나타내는 개체명으로 태깅할 수 있으나 ‘한국인’이라는 단어가 보통명사로 쓰이고 있으므로 개체명이 될 수 없다.

[예외] 드라마, 영화의 제목처럼 하나의 덩어리를 이루는 경우

[예외] FTA, 올림픽, 월드컵 등 협정, 이벤트를 나타내는 경우

(2) 개체명은 최장일치를 기본으로 한다.

길이가 짧은 개체명이 긴 개체명에 포함될 수 있는 경우, 길이가 긴 개체명을 우선시한다. 예를 들어, ‘미국 컬럼비아 대학교’의 경우 ‘미국’이라는 위치(LOC)를 나타내는 개체명과 ‘컬럼비아 대학교’라는 교육기관(ORG)를 나타내는 개체명으로 각각 태깅할 수 있지만 ‘미국 컬럼비아 대학교’를 모두 묶어 기관(ORG)으로 태깅한다. 동일한 태그가 연속하여 나타날 때, 연속된 개체명을 하나로 묶을 수 있을 경우 하나의 개체명으로 태깅한다. (ex. ‘부산 사직구장’)

[예외] 주소. (3)번 항목 참고

[예외] ‘14타수 7안타’는 전체를 수량표현으로 묶을 수 있으나, ‘14타수’와 ‘7안타’는 수량의 단위가 다르다. 이 경우에는 ‘14타수’와 ‘7안타’로 분리하여 태깅한다.

(3) 개체명이 특정 개념, 사실, 대상을 구체화하더라도 주소와 같이 각 단어가 개체명이 될 수 있을 경우에는 각 단어를 개별의 개체명으로 태깅한다.

‘부산광역시 영도구 태종로 727’의 경우, ‘부산광역시’ ‘영도구’ ‘태종로 727’로 태깅한다. ‘태종로’와 ‘727’를 함께 묶어서 태깅하는 경우는 뒤의 숫자가 태종로의 특정 장소를 구체화하기 때문이다.

(4) 특정한 개념, 사실, 대상을 구체적으로 가리키지 못할 경우에는 태깅하지 않는다.

‘몇 번’, ‘몇 회’, ‘오늘’, ‘가을’, ‘며칠 뒤’ 등처럼 불특정한 개념이나 상대적인 개념은 태깅하지 않는다.

[예외] 상대적인 개념이 수량 표현과 관련된 경우 ‘기타 수량표현’을 의미하는 NOH로 태깅한다.

(5) 하나의 개체명이 여러 개의 태그가 가능할 경우, 문맥을 통해 가장 알맞은 태그로 태깅한다.

2. 한국어 개체명 태그의 정의

가. 개체명(ENAMEX)

개체명은 주로 고유명사에 해당하며 아래와 같은 형태가 있다.

1) 인명(person) : <ne tag="PER"> </ne>

사람의 성씨(姓氏)나 이름(음차한 외국인의 이름, 영어, 한자, 일본어 등으로 표현된 외국인의 이름도 포함)에 해당한다. 그 밖에도 생물(生物)을 부르기 위에 붙인 이름, 별명으로 사용되는 말도 포함한다. 실존 인물, 신화/종교에 등장하는 인물, 만화/게임/소설/드라마 등에 등장하는 인물이나 존재에 대한 이름도 여기에 해당한다.

[참고] 직책, 존칭 등을 표현하는 말은 PER에 포함시키지 않는다.

[참고] 볼 빨간 사춘기, 방탄소년단, 트와이스처럼 2명 이상이 모여 그룹으로 활동하는 경우는 ORG로 태깅한다.

[참고] 1인 기업, 1인 그룹 등 기업이나 그룹인 경우에도 ORG로 태깅한다.

[참고] ‘홍련왕’이나 ‘마이다스의 손’, ‘될성부른 떡잎’, ‘국민 여동생’처럼 관습적으로 사용되는 표현을 별명으로 사용한 경우는 태깅하지 않는다.

- 조선시대 풍속화의 대가로 손꼽히는 ‘<ne tag="PER">단원 김홍도</ne>’와 ‘<ne tag="PER">혜원 신윤복</ne>’은 동시대에 활동하며, 수많은 그림을 화첩(畫帖)으로 남겼다.
- <ne tag="PER">박</ne> 감독은 <신세계> <브이아이피> 등 한국형 장르·누아르물을 연출해왔다.
- <ne tag="PER">환웅</ne>이 천하에 뜻을 두어 인간 세계를 탐내거늘...
- 중국의 소설 『서유기』에 등장하는 원숭이의 이름. 화과산의 돌에서 태어나 수렴동에서 <ne tag="PER">미후왕</ne>이라고 칭하고, 선인으로부터 72반의 선술(仙術)을 배워서 신통력을 몸에 익히고 <ne tag="PER">손오공</ne>이라는 이름을 받았다.
- 우리집 고양이 <ne tag="PER">똥똥이</ne>는 어미인 <ne tag="PER">행운이</ne>처럼 우리 엄마 뒤편 졸졸 쫓아다닌다.
- 그 선수는 <ne tag="PER">한국산 핵잠수함</ne>이라는 절명으로 불리고 있다.

2) 기관/조직(organization) : <ne tag="ORG"> </ne>

여러 사람이 속해서 이루어진 공동체로서 특정 집단을 지칭하는 단어이다. 회사명(google), 정부조직명(과학기술정보통신부), 위원회(법사위원회, 올림픽조직위원회), 교육기관(한국해양대학교), 2명 이상의 구성원으로 활동하는 단체(볼 빨간 사춘기, 트와이스) 등이 기관/조직 범주에 속한다.

[참고] 특정 단체의 별명도 ORG로 태깅한다. (ex. '한화 이글스'의 별명인 '독수리 군단')

[참고] 2018 러시아 월드컵 '예선 G조', '예선 F조' 등과 같이 여러 명이 속하여 일정 기간 이상 함께하는 경우에는 ORG로 태깅한다.

[참고] '미국 컬럼비아 대학교'와 같이 국가/지역/장소명이 고유 명칭에 포함되어 있는 경우에는 이를 전부 ORG로 태깅한다. ⇒ 최장일치

[참고] '한·미', '한미' 등과 같이 2개 이상의 기관을 지칭하는 단어는 덩어리째로 ORG로 태깅한다.

[참고] 정부, 협회, 연구기관, 대검찰청, 대법원, 헌법 재판소처럼 일반명사에 해당하는 표현이 단독으로 쓰인 경우는 태깅하지 않는다.

[참고] '문재인 정부'의 경우 '문재인'이라는 인명(PER)에 해당하는 개체명이 '정부'라는 일반명사를 수식하는 형태이므로 ORG로 태깅하지 않는다.

단, '미국방부'의 경우 기관범주(ORG)+기관범주(ORG)=하나의 기관(ORG)으로 성립하므로 ORG로 태깅한다.

- '<ne tag="ORG">상륙수부대</ne>'는 당초 <ne tag="ORG">UN</ne>에 통보한대로 올 7월까지 소말리아에서 근무하고 철수한다는 것이 기본 방침이다.
- 문재인 대통령이 11일 "<ne tag="ORG">주한미군</ne>은 <ne tag="ORG">한미동맹</ne>의 문제이지 <ne tag="ORG">북미</ne> 간 비핵화 협상 과정에서 논의될 의제가 아니다"라며 주한미군 철수 가능성에 선을 그었다.
- <ne tag="ORG">일본</ne>은 2009년 제2회 대회부터는 적극적으로 대회 운영의 개선을 요구할 방침이다.
- <ne tag="ORG">일본</ne>·<ne tag="ORG">미국</ne>·<ne tag="ORG">프랑스</ne>·<ne tag="ORG">노르웨이</ne>의 4개국은 제2회 대회부터는 적극적으로 대회 운영의 개선을 요구할 방침이다.

3) 장소/위치(location) : <ne tag="LOC"> </ne>

(정부의 역할이 아닌 장소를 나타낼 때의) 국가명, 지역명칭과 행정 구역을 나타내는 단어가 여기에 속한다. 국가명이나 행정구역과는 별개의 산/강 같은 자연에 대한 고유명칭(팔공산, 낙동강, 금호강)이나 관광명소, 경기장 등 특정 위치를 나타낼 수 있는 단어이다.

[참고] 상대적인 위치에 해당하는 개념은 태깅하지 않는다.

[참고] 한류, 방한, 내한, 방미 등과 같이 LOC에 해당하는 단어가 다른 단어와 결합되어 복합명사로 굳어진 경우에는 태깅하지 않는다.

* ‘한국영화’ 등과 같은 경우에는 지명에 해당되는 부분을 LOC로 태깅한다.

[참고] ‘한국야구’, ‘한국 프로축구’, ‘한국 농구’ -> ORG

[참고] 한국인처럼 한국(고유명사)+인(접사)의 형태로 분리가 될 경우에도 태깅하지 않는다. (‘:’ 하나의 단어로 굳어진 형태), [태국산, 한국산 등도 여기에 속함]

- 북한은 <ne tag="LOC">싱가포르</ne>에서 열린 미국과의 회담에서 북한 핵시설에 대한 국제사찰 문제 등을 논의했다.
- <ne tag="LOC">서울역</ne>, <ne tag="LOC">김포공항</ne>의 표가 금일 오전에 대부분 매진됐다.
- 문 대통령, 모디 총리, 이재용 부회장과 <ne tag="LOC">삼성전자 제2공장</ne> 준공식 참석
- <ne tag="LOC">부산광역시</ne> <ne tag="LOC">영도구</ne> <ne tag="LOC">태종로 727</ne>

4) 기타 고유명사 : <ne tag="POH"> </ne>

제품명, 약품명, 특정 경기를 지칭하는 표현이나 행사, 회의, 사건 등을 나타내는 표현이다. 영화 제목, 노래 제목, 드라마 제목 등도 여기에 속한다. TV 채널 이름이나 일간지, 주간지 등도 POH로 태깅한다. 이메일 주소, 전화번호도 하나의 고유한 개체에 해당하므로 POH로 태깅한다.

[참고] ‘2018 러시아 월드컵’의 경우 2018은 날짜(DAT)로 태깅하고, ‘러시아 월드컵’ 부분은 POH로 태깅한다.

[참고] FTA, 월드컵, 올림픽, 영화제, 대회와 같이 일반명사가 단독으로 쓰인 경우에는 태깅하지 않는다.

[참고] ‘슈렉2, 3’처럼 동일한 작품의 시리즈가 열거 되어 나오는 경우, 제목 부분(슈렉)만 POH로 태깅하고 시리즈에 해당되는 번호(2,3)은 NOH로 태깅한다. 단, ‘슈렉 3’처럼 단독으로 쓰인 경우에는 전체를 POH로 태깅한다.

- ‘<ne tag="POH">이리와 안아줘</ne>’ 윤지혜, 김서형 살인사건 파헤쳤다.
- 차명석은 <ne tag="POH">MBC ESPN</ne>의 해설위원과 LG트윈스 투수코치로도 활동했다.
- 검찰은 <ne tag="POH">아시아나 기내식 대란</ne>과 관련하여 시민단체가 박상구 금호아시아나그룹 회장을 배임 혐의로 고발한 사건에 대해 수사에 착수했다.
- ‘<ne tag="POH">신과함께-인과 연</ne>’韓영화 최초 전세계 IMAX 스크린 상영 확정
- 이번 월드컵에 참가하는 프랑스는 ‘황금세대’라고 불린다.
*어떤 월드컵인지 알 수 있는 방법이 없으므로 태깅하지 않는다.
- 제 이메일 주소는 <ne tag="POH">kmounlp@naver.com</ne> 입니다.
- <ne tag="POH">슈렉</ne> 2, 3을 볼래? 아니면 <ne tag="POH">해리포터와 죽음의 성물 part 1</ne>을 볼래?

나. 시간표현(TIMEX)

시간표현은 **절대적인 시간표현**을 기준으로 한다. 절대적인 시간표현이란 ‘2002년 12월 25일’과 같이 구체적인 시간을 나타내는 표현이며, 상대적인 시간표현은 ‘오늘’, ‘어제’, ‘올해’, ‘금년’, ‘이틀 뒤’ 등과 같이 특정 기준이 있어야 정확한 시점을 찾을 수 있는 표현이다.

특정 시기에 한 번씩을 의미하는 ‘마다’가 붙은 기간 표현, 특정 현상이 한 번 나타나고 다음번 되풀이 되기까지에 해당되는 ‘-주기’, 일 년을 단위로 돌아오는 돌을 세는 단위 ‘-주년’, 나이나 햇수의 차례를 의미하는 ‘-연차/-년차’, 앞말이 가리키는 동안이나 거리 등을 의미하는 ‘-만’과 결합된 표현은 시간표현으로 보지 않는다.

5) 날짜(date) : <ne tag="DAT"> </ne>

절대적인 년, 월, 일에 관련된 표현이다. 기념일(한글날, 광복절, 현충일 등), 절기(節氣)처럼 특정 날짜를 지칭하는 표현도 여기에 속한다. 일요일, 월요일, 목요일 등 요일도 날짜에 속한다.

[참고] ‘46일간’, ‘12개월’, ‘징역 8월’, ‘5년차 선수’에서 46일, 12개월, 8월, 5년은 날짜로 볼 수 없다.

위의 예시에 나타난 표현은 (상대적인)기간, 연차에 해당하는 개념으로 DAT로 태깅하지 않는다.

[참고] ‘6월 한 달 간’처럼 특정 기준 날짜와 상대적인 기간이 함께 등장하는 경우에는, 절대적인 날짜에 해당되는 부분만 DAT로 태깅한다.

[참고] ‘오늘’, ‘어제’, ‘올해’ 등과 같은 상대적인 표현은 태깅하지 않는다.

[참고] ‘7월 말에 출시된다’에서 ‘7월 말’처럼 상대적인 날짜 개념과 함께 쓰인 경우 절대적인 날짜에 해당되는 부분만 DAT로 태깅한다.

[참고] 기간(duration)을 나타내는 표현 ‘부터’, ‘까지’와 함께 쓰이는 경우 다음 원칙을 따른다.

(1) 2016년부터 2018년까지 계약을 유지하기로 했다.

*2016년부터 2018년은 기간에 해당하는 표현이므로 DAT로 태깅하지 않는다.

(2) 그는 2010년부터 매해 3천만원을 재단에 기부해왔다.

*기간의 시작을 나타내는 부터를 제외한 날짜부분만 DAT로 태깅한다.

(3) 이 계약은 2018년까지 유효합니다.

*기간의 끝을 나타내는 까지를 제외한 날짜부분만 DAT로 태깅한다.

- <ne tag="DAT">4월 20일</ne>(<ne tag="DAT">금요일</ne>)은 봄의 마지막 절기인 <ne tag="DAT">곡우</ne>(<ne tag="DAT">穀雨</ne>)다.

- 일본은 <ne tag="DAT">2009년</ne> 제2회 대회부터는 적극적으로 대회 운영의 개선을 요구할 방침이다.

- 리오스는 지난 시즌 17승으로 공동 다승왕에 오르는 등 <ne tag="DAT">2002년</ne>부터 지난 시즌까지 줄곧 10승 이상씩을 올리고 있다.

- 그는 <ne tag="DAT">2000년</ne>부터 시애틀 매리너스 마무리로 활약하며 <ne tag="DAT">2003년</ne>까지 4년간 7승 16패 129세이브를 기록했다.

6) 시간(time) : <ne tag="TIM"> </ne>

절대적인 시, 분, 초에 관련된 표현이다. ‘정오’는 12시, ‘자정’은 00시니까 태깅한다. ‘오전’, ‘오후’ 등은 상대적인 시간에 해당한다.

[참고] ‘7시간’, ‘경기 시간 120분 동안’, 등에서 7시, 60분은 시간으로 볼 수 없다.

위의 예시에 나타난 표현은 (상대적인)기간에 해당하는 개념으로 태깅하지 않는다.

[참고] ‘9시 반 이후’처럼 특정 기준 시간과 상대적인 표현이 함께 등장하는 경우에는, 절대적인 시간에 해당되는 부분만 TIM으로 태깅한다.

[참고] ‘오전’, ‘오후’, ‘저녁’, ‘밤’ 등은 단독으로 쓰일 경우 상대적인 시간으로 취급하여 태깅하지 않는다.

단, ‘오전 7시’처럼 절대적인 시간과 함께 쓰이면 시간의 의미를 구체화시켜주는 역할을 수행하므로 함께 TIM으로 태깅한다.

[참고] ‘정오’는 낮 12시를 ‘자정’은 밤 12시를 의미하므로 단독으로 쓰여도 TIM으로 태깅한다.

[참고] 기간(duration)을 나타내는 표현 ‘부터’, ‘까지’와 함께 쓰이는 경우 다음 원칙을 따른다.

(1) 7시부터 10시까지 가야금 수업을 듣는다.

*7시부터 10시는 기간에 해당하는 표현이므로 TIM으로 태깅하지 않는다.

(2) 10시 35분부터 5분마다 알람이 울려서 무슨 일인가 했다.

*기간의 시작을 나타내는 부터를 제외한 시간부분만 TIM로 태깅한다.

(3) 이 계약은 18:00까지 유지된다.

*기간의 끝을 나타내는 까지를 제외한 시간부분만 TIM로 태깅한다.

- 목격자는 “<ne tag="TIM">10시 35분</ne>부터 5분마다 알람이 울려서 무슨 일인가 했어요”라고 말했다.

- ‘달려라 방탄’은 11일 <ne tag="TIM">오후 6시</ne>를 시작으로 8월 29일까지 8주간 매주 수요일 <ne tag="TIM">오후 6시</ne>에 방영된다.

- 또봇V의 애니메이션은 오는 7월 12일(목) <ne tag="TIM">오후 5시 15분</ne>에 KBS 2TV를 통해 첫 방송된다.

7) 기간(duration) : <ne tag="DUR"> </ne>

어떤 일이 지속되는 기간을 표현하는 말이다. ‘백악기’, ‘쥐라기’와 같이 학문적으로 정의가 된 기간은 duration으로 태깅한다.

[참고] ‘7시부터 10시까지 가야금 수업을 듣는다.’라는 문장에서 기간에 해당되는 표현은 ‘7시부터 10시까지’이다. 이 표현에서 실제로 DUR로 태깅되는 부분은 ‘7시부터 10시’이다. (‘:’ 조사 ‘-까지’를 제외)

[참고] ‘백악기’, ‘쥐라기’와 같이 학자들이 정의해놓은 시기는 기간으로 태깅한다.

[참고] 1980년대와 같이 -년대는 1980년~1989년을 나타내므로 기간으로 태깅한다.

[참고] 단순히 고대, 중세, 현대와 같이 표기한 경우는 개체명으로 태깅하지 않는다.

[참고] 숫자 + 주기/연간/연(년)도/학년도 등의 기간을 나타내는 단위와 결합된 표현은 DUR로 태깅한다.

특정 현상이 한 번 나타나고 다음번 되풀이되기까지 기간에 해당되는 ‘-주기’
한 해 동안을 나타내는 ‘-연간’
일정한 기간 단위로서의 그해를 나타내는 ‘-연도/-년도’
한 학년의 과정을 배우는 기간을 의미하는 ‘-학년도’
등...

- 꼴찌는 평화협정이 체결돼 겨우 <ne tag="DUR">16 년간</ne>의 내전에서 벗어나게 된 아프리카의 모잠비크로 1인당 소득은 단돈 60달러에 불과했다.
- 한국의 틱스피리트 훈련 중단 발표가 미국의 3단계 고위급회담 발표보다 <ne tag="DUR">14시간</ne> 먼저 있게 되는 것이다.
- 서울 종로구 낙원동에 위치한 서울아트시네마(구 허리우드 극장)는 <ne tag="DUR">29-7월7일</ne> 한국과 일본, 미국, 유럽의 청춘영화를 모아 상영한다.
- 그 영화는 <ne tag="DUR">6월 4일부터 30일</ne>까지 상영된다.

다. 수량표현(NUMEX)

수량표현은 숫자와 단위를 함께 사용한 표현이나 숫자를 단독으로 사용한 경우를 나타내며 금액, 비율, 기타 숫자표현의 형태로 태깅할 수 있다.

1) 금액(money) : `<ne tag="MNY"> </ne>`

돈의 양을 나타내는 표현이다. (화폐단위 한정, 비트코인 이런 거 안됨...)

[참고] 돈의 양을 나타내는 단위와 함께 태깅한다.

[참고] '최소 2만원 ~ 최대 50만원'이라는 표현처럼 범위를 나타내는 ~ 혹은 -와 함께 쓰일 경우 ~, -를 제외한 실제 돈의 액수를 나타내는 표현만 태깅한다.

[참고] 열거하기 위한 ,를 함께 사용한 경우 ,를 제외하고 태깅한다.

[참고] 몇 만원처럼 불특정한 경우는 태깅하지 않는다.

[참고] 수십만원 등처럼 대략적으로 유추 가능한 경우에는 금액으로 태깅한다.

- 그녀는 한 해에 대략 `<ne tag="MNY">8만 파운드</ne>` 가량 번다.
- 92년 현재 토지비용을 제외한 건설물량이 약 `<ne tag="MNY">47조원</ne>`으로 GNP의 31.5%에 달한다.
- 매월 최소 `<ne tag="MNY">2만원</ne>` ~ 최대 `<ne tag="MNY">50만원</ne>`까지 청약이 가능하다.
- 몇 만원인지 정확히 알 수 없지만 최소 `<ne tag="MNY">수십여만원</ne>`으로 추측된다.

2) 비율(rate) : `<ne tag="PNT"> </ne>`

백분율, 혹은 다른 수나 양에 대한 어떤 수나 양의 비(比)에 해당하는 표현이다.

(ex. %, %p, 할푼리 등)

[참고] 단위와 함께 태깅한다.

[참고] 범위를 나타내는 ~ 혹은 -와 함께 쓰일 경우 ~, -를 제외한 실제 비율을 나타내는 표현만 태깅한다.

[참고] 열거하기 위한 ,를 함께 사용한 경우 ,를 제외하고 태깅한다.

[참고] 소수점이라도 비율에 해당하지 않으면 PNT로 태깅하지 않는다.

(ex. 2.5명, 3.4게임, 36.5도 등...)

- 92년 현재 토지비용을 제외한 건설물량이 약 47조원으로 GNP의 `<ne tag="PNT">31.5%</ne>`에 달한다.
- 구자욱이 타율 `<ne tag="PNT">3할 6푼 4리</ne>`를 기록했고 최형우는 타율 `<ne tag="PNT">3할 6푼리</ne>`를 기록했다.

2) 기타 숫자표현 : <ne tag="NOH"> </ne>

특정한 단위를 가지고 나이, 면적, 거리, 속도, 온도, 부피, 무게, 에너지 등의 양이나 정도를 표현하는 말이다. 단위성 의존명사와 함께 사용되는 경우이다. 이 외에도 개체의 양이나 수를 표현하거나 수 자체를 표현하는 계수(係數), 사물의 순서를 나타내는 수인 서수(序數) 표현의 일부가 여기에 속한다.

(※ 개체명에 포함되지 않는 경우의 수가 많으나, 자동번역에 활용하기 위해 추가한 항목이다.)

[참고] 단위와 함께 태깅한다.

[참고] 범위를 나타내는 ~ 혹은 -와 함께 쓰일 경우 ~, -를 제외하고 태깅한다.

단, -가 점수를 의미할 경우에는 함께 태깅한다. (ex. 5-4)

[참고] 서수(序數)의 경우 '제26회', '제5차년도' 등과 같은 표현은 NOH로 태깅한다.

단, '첫째', '둘째' 등과 같은 관형사는 태깅하지 않는다.

[참고] 열거하기 위한 ,를 함께 사용한 경우 ,를 제외하고 태깅한다.

[참고] 단위가 다른 수량표현이 연속으로 등장할 경우, 단위 별로 NOH로 태깅한다.

[참고] 숫자 + 마다/주년/연(년)차/만 등의 단위와 결합된 표현은 NOH로 태깅한다.

특정 시기에 한 번씩을 의미하는 '마다'
특정 현상이 한 번 나타나고 다음번 되풀이되기까지 기간에 해당되는 '-주기'
일 년을 단위로 돌아오는 돌을 세는 단위 '-주년'
나이나 햇수의 차이를 의미하는 '-연차/-년차'
앞말이 가리키는 동안이나 거리 등을 의미하는 '-만(에)'
등...

- 올해 출시 <ne tag="NOH">30주년</ne>을 맞은 오리온의 '포카칩'이 국내 누적매출액 1조4천억원을 돌파했다.
- 대한기계설비단체총연합회(회장 백종윤)는 12일 서울 여의도 63컨벤션센터 그랜드볼룸에서 <ne tag="NOH">600여명</ne>의 기계설비인들이 참석한 가운데 <ne tag="NOH">제3회</ne> 기계설비의날 기념식(사진)을 개최했다.
- 채은성은 4일 서울 잠실야구장에서 열린 NC와의 경기에 선발 <ne tag="NOH">5번</ne> 겸 우익수로 나와 <ne tag="NOH">4회</ne> <ne tag="NOH">2사</ne> <ne tag="NOH">1</ne>, <ne tag="NOH">2루</ne>에서 홈런을 쳐냈다. <ne tag="NOH">6-2</ne>, <ne tag="NOH">넉 점</ne> 차이로 앞서고 있던....
- 이 후보는 기자회견에서 경력 <ne tag="NOH">20년차</ne> 이상 교사에게 유급 장기 휴가를 부여하는 유급...
- <ne tag="NOH">양</ne>팀은 각각 이정식, 송산이라는 백업 포수가 있으나 주전들과 기량차가 현격해 이들의 빈자리를 얼마만큼 채워줄 수 있을지는 미지수다.
- 슈렉 <ne tag="NOH">2</ne>, <ne tag="NOH">3</ne>을 볼래? 아니면 해리포터와 죽음의 성물 part 1을 볼래?

(※ 야구용어에서 '만루'는 '루가 모두 채워졌다'라는 의미이므로 태깅하지 않는다.)

(※ 양대 산맥에서 양대의 양은 태깅하지 않는다. [∵ 양대라는 보통 명사가 존재])