

국립국어원 개체명 분석 말뭉치

(버전 1.0)

- **자료명:** 국립국어원 개체명 분석 말뭉치
- **공개일**
 - (버전 1.0) 2020. 8. 25.
- **자료 유형:** 텍스트
- **관련 사업:** 개체명 분석 말뭉치 구축(2019)
- **자료 설명**
 - **내용**
 - 개체명의 경계를 인식하고 15개 의미 분류 체계에 따른 표지를 부착한 말뭉치임. 한국전자통신연구원(ETRI)의 ‘세부분류 개체명 가이드라인 2018’ 지침에 준하여 분석하였음.
 - **분석 표지:** 15개
 - PERSON(PS), LOCATION(LC), ORGANIZATION(OG), ARTIFACTS(AF), DATE(DT), TIME(TI), CIVILIZATION(CV), ANIMAL(AM), PLANT(PT), QUANTITY(QT), STUDY_FIELD(FD), THEORY(TR), EVENT(EV), MATERIAL(MT), TERM(TM)
 - **분량**
 - 총 300만 어절(문어 200만 어절, 구어 100만 어절)
 - **파일 형식:** JSON(UTF-8 인코딩)
 - **파일 수 및 크기:** 파일 2개, 총 909MB

· 파일 명명 규칙

자리	1	2	3	4	5	6	7	8	9	10	11	12	13	14
속성	유형	구성 방법	주석 단계	구축 연도	일련번호(8자리)									
정의 값	N: 신문 S: 구어	X: 추출	NE: 개체명	19: 2019년	00000001 ~ 99999999 (여덟 자리 일련번호)									
※ 예시: NXNE1902008030.json 2019년에 구축한 신문 자료 가공 개체명 말뭉치 파일 SXNE1902007240.json 2019년에 구축한 구어 자료 가공 개체명 말뭉치 파일														

· 인용:

(국문) 국립국어원(2020). 국립국어원 개체명 분석 말뭉치(버전 1.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2020). NIKL Named Entity Corpus (v.1.0). URL: <https://kli.korean.go.kr/corpus>

· 예시

```
{
  "id": "NWRW1800000029.315.2.3",
  "form": "멕시코 파키스탄 방글라데시 베트남 역시 해외파견 근로자의 송금이 한몫을
한다.",
  "word": [
    {
      "id": 1,
      "form": "멕시코",
      "begin": 0,
      "end": 3
    },
    {
      "id": 2,
      "form": "파키스탄",
      "begin": 4,
      "end": 8
    },
    {
      "id": 3,
      "form": "방글라데시",
      "begin": 9,
      "end": 14
    },
    {
      "id": 4,
      "form": "베트남",
      "begin": 15,
      "end": 18
    },
    {
      "id": 5,
      "form": "역시",
      "begin": 19,

```

```

        "end": 21
    },
    {
        "id": 6,
        "form": "해외파견",
        "begin": 22,
        "end": 26
    },
    {
        "id": 7,
        "form": "근로자의",
        "begin": 27,
        "end": 31
    },
    {
        "id": 8,
        "form": "송금이",
        "begin": 32,
        "end": 35
    },
    {
        "id": 9,
        "form": "한몫을",
        "begin": 36,
        "end": 39
    },
    {
        "id": 10,
        "form": "한다.",
        "begin": 40,
        "end": 43
    }
},
"NE": [
    {
        "id": 1,
        "form": "멕시코",
        "label": "LC",
        "begin": 0,
        "end": 3
    },
    {
        "id": 2,
        "form": "파키스탄",
        "label": "LC",
        "begin": 4,
        "end": 8
    },
    {
        "id": 3,
        "form": "방글라데시",
        "label": "LC",
        "begin": 9,
        "end": 14
    },
    {

```

```
        "id": 4,  
        "form": "베트남",  
        "label": "LC",  
        "begin": 15,  
        "end": 18  
    },  
    {  
        "id": 5,  
        "form": "근로자",  
        "label": "CV",  
        "begin": 27,  
        "end": 30  
    }  
  ]  
},
```

· 자료 내용 문의: 02-2669-9638