

국립국어원 감성 분석 말뭉치 2020 (버전 1.0)

- **자료명:** 국립국어원 감성 분석 말뭉치 2020
- **공개일**
 - (버전 1.0) 2021. 8. 19.
- **자료 유형:** 텍스트
- **관련 사업:** 말뭉치 감성 분석 및 연구(2020)
- **자료 설명**
 - **내용**
 - '19년 웹 말뭉치에서 선정된 후기 문서 2,081건을 대상으로 작성자의 주관성이 드러나는 표현 (감성 표현)을 대상으로 감성 분석 정보를 부착한 말뭉치

(감성 점수 부착 기준) * 문서별·표현별 부착

-2	-1	0	1	2
강한 부정	부정	중립	긍정	강한 긍정

(주제 및 표현 분류)

주제 범주(Subject_category)	제품, 영화, 여행
주제(Subject)	감성 표현이 가리키는 구체적 대상 예) 닥터지마스크팩, 과속스캔들, 남이섬 등
표현 범주(Expression_category)	감성 표현의 유형 분류 예) 기능/성능(제품), 감독/스텝(영화), 숙소(여행) 등

- **분량**
 - 문서 총 2,081건

	제품	영화	여행	합계
블로그·게시판	39	181	53	273
누리소통망(SNS)	1,517	206	85	1,808
합계	1,556	387	138	2,081

· 파일 형식: JSON(UTF-8 인코딩)

· 파일 수 및 크기: 파일 1개, 총 2MB(ZIP 압축)

· 인용:

(국문) 국립국어원(2021). 국립국어원 감성 분석 말뭉치(버전 1.0). URL: <https://kli.korean.go.kr/corpus>

(영문) National Institute of Korean Language (2021). NIKL Sentiment Analysis Corpus (v.1.0).

URL: <https://kli.korean.go.kr/corpus>

· 파일 명명 규칙

자리	1	2	3	4	5	6	7	8	9	10	11	12	13	14
속성	매체	장르	주석 단계		구축 연도		일련번호(8자리)							
정의값	E: 웹	X: 추출	SA: 감성 분석 말뭉치		20: 2020년		00000001 ~ 99999999 (여덟 자리 일련번호)							
※ 예시: EXSA2002108040.json 2020년에 구축한 웹 원시 말뭉치 추출 감성 분석 말뭉치														

· 예시

```
{
  "id": "EXSA2002108040",
  "metadata": {
    "title": "국립국어원 웹 말뭉치 추출 EXSA2002108040",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2020",
    "category": [
      "웹>리뷰>누리소통망",
      "웹>리뷰>블로그"
    ],
    "annotation_level": [
      "감성 분석"
    ],
    "sampling": "부분 추출 - 특정 부분 추출"
  },
  "document": [
    {
      "document_id": "ERRW1903002753.6",
      "document_score": 1,
      "metadata": {
        "title": "반신욕 한약재 티백",
        "author": "June",
        "publisher": "egloos"
      }
    }
  ]
}
```

```

        "date": "20090524",
        "url": "http://kimmari02.egloos.com/2389745"
    },
    "paragraph": [
        {
            "paragraph_id": "ERRW1903002753.6.1",
            "paragraph_form": "나이가 xx대 줄에 들어서자 전부터 그랬지만 더욱 건강
에 매진하는 나. 주변의 친구들도 하나 둘 그 길을 걷기 시작하고, 걸보기에는 너무 건강한 그녀로부터 목욕
용 한약재 티백을 쓰고 있다는 소릴 듣는다. 나 또한 생리통이 심하면 만사 다 집어치고 집에 가고 싶어지기
때문에 좋다는 전언에 따라 구입했다. 익모초+약쑥 티백 24개 들어.. ㅋㅋ"
        },
        {
            "paragraph_id": "ERRW1903002753.6.2",
            "paragraph_form": "깔끔하게 부직포 포장으로 되어 있어서 그냥 뜨거운 물
에 풍덩 넣어놓고 좀 휘젓어주면 금방 우려난다. 목욕할 때마다 넣어봤는데(샤워는 자주 해도 목욕은 그렇게
자주가 아님.. 이것도 약재는 약재이므로 용법은 알아서;) 신선한 한약풀 냄새가 욕실에 퍼져서 기분이 좋아졌
다. 아직 때가 안 되서 효과까지는 모르겠는데 가려운 피부에도 효과가 있었으면 좋겠네. 박하 같은 것도 팔
던데 지금으로서는 대만족이라 다음에는 상쾌하게 박하 사고 싶다. 혹시 오래된 거 팔지 않나 고민했었는데
쑥향기 자체가 페퍼민트처럼 신선하고 포장도 깔끔하고 사용도 간편하고 참.. 우리나라 인터넷 시장도 좋은
거 같다. 주문하니 이렇게 물에 넣기만 하면 되게 딱딱 만들어서 집까지 숨 배달해주고.."
        },
        [...]
    ],
    "sentiment_expression": [
        {
            "expression_id": 1,
            "expression": [
                {
                    "expression_form": "깔끔하게 부직포 포장으로 되어 있어서",
                    "paragraph_id": "ERRW1903002753.6.2",
                    "begin": 0,
                    "end": 20
                }
            ],
            "expression_score": 1,
            "expression_category": [
                "디자인"
            ],
            "subject_category": "제품",
            "subject": [
                "목욕용 한약재 티백"
            ]
        },
        {
            "expression_id": 2,
            "expression": [
                {
                    "expression_form": "금방 우려난다.",
                    "paragraph_id": "ERRW1903002753.6.2",
                    "begin": 47,
                    "end": 55
                }
            ],
            "expression_score": 1,
            "expression_category": [
                "기능/성능"
            ]
        }
    ],

```

```

        "subject_category": "제품",
        "subject": [
            "목욕용 한약재 티백"
        ]
    },
[...]
· 하나의 표현이 두 개 이상의 문단에 걸쳐 나타나는 경우 아래와 같이 처리함 (예 “저 따위 소리를 하는
  건”)
    {
        "expression_id": 13,
        "expression": [
            {
                "expression_form": "저 따위 소리를",
                "paragraph_id": "ERRW1903002756.49.17",
                "begin": 48,
                "end": 56
            },
            {
                "expression_form": "하는 건",
                "paragraph_id": "ERRW1903002756.49.18",
                "begin": 0,
                "end": 4
            }
        ],
        "expression_score": -1,
        "expression_category": [
            "관람경험"
        ],
        "subject_category": "영화",
        "subject": [
            "추격자"
        ]
    },

```

· 자료 내용 문의: 02-2669-9638