# 한국어 신조어 말뭉치 구축 및 신조어 중요도 측정 방법에 대한 연구

김현지○, 정상근*, 황태욱
충남대학교 컴퓨터융합학부
hjkim2714@gmail.com, hugman@cnu.ac.kr, taewook5295@gmail.com

# A Study of the construct Korean New Word Corpus and Metric of New Word Importance

Hyunji Kim○, Sangkeun Jung*, Taewook Hwang
Department of Computer Science and Engineering, Chungnam National University

**요 약**

신조어는 자연어처리에 있어 대단히 중요하며, 시스템의 전체 성능에 직접적인 영향을 미친다. 일단위, 주단위로 신규 발생하는 어휘들에 대해, 자동으로 신규성 및 중요도가 측정되어 제공된다면, 자연어처리 연구 및 상용시스템 개발에 큰 도움이 될 것이다. 이를 위해, 본 연구는 한국어 말뭉치 KorNewVocab을 새로이 제시한다. 먼저, 신조어가 가져야 할 세부 중요 조건을 1)신규 어휘 2)인기 어휘 3)지속 사용 어휘로 정의하고, 이 조건을 만족하는 신조어 말뭉치를 2019.01~2019.08까지의 뉴스기사를 중심으로 신조어 412개와 4,532 문장으로 구성된 신조어 말뭉치를 구축하였다. 또한, 본 말뭉치의 구축에 활용된 반자동 신규어휘 검출 및 중요도 측정 방법에 대해 소개한다.

주제어: 신조어, 데이터셋, 자연어 처리 연구

## 1. 서론

최근 인터넷을 통한 소통의 장이 발달함에 따라, 사람들의 자유롭고 즉각적인 반응이 유도되어 이어지고 있다. 소셜 네트워크 서비스(SNS)와 인터넷 사이트 등을 통한 자유로운 소통과 교류의 가장 큰 특징으로 새로운 말들이 지속적으로 생성되고 있으며 이를 신조어라 한다[1].

신조어는 데이터에 의해 성능이 좌우되는 자연어 처리 연구에 영향을 주게 된다. 여기서 자연어(Natural language)란 인간이 일상생활에서 사용하는 언어를 말하며, 자연어 처리는 자연어의 의미를 분석하여 컴퓨터가 처리할 수 있도록 하는 일을 의미한다.

신조어가 데이터셋에 포함될 경우, 자연어 처리 모델 성능 평가에 부정적인 영향을 끼치게 되며, 이를 해결하기 위한 많은 연구가 진행되고 있다.

외국의 경우, 신조어로 인한 성능 저하를 막기 위하여 문맥 표현과 대규모 어휘 배경지식을 활용한 신조어의 슬롯 태깅 모델 개발[2] 및 신조어와 의미가 유사한 사전 속 단어를 찾아 대체하는[3] 등 학습 데이터셋에서 존재하지 않는 신조어를 처리하기 위한 많은 과제들이 활발히 연구되고 있다.

국내의 경우 한국어와 관련된 연구를 진행할 시, 한국어 분석을 위한 형태소 분석기 사용이 필수적이다. 하지만, 현재 사용되고 있는 형태소 분석기는 지속적인 업데이트를 제공해주지 않는 이상, 과거 데이터를 기반으로 개발되어 신조어 대응 성능이 떨어질 수밖에 없다.

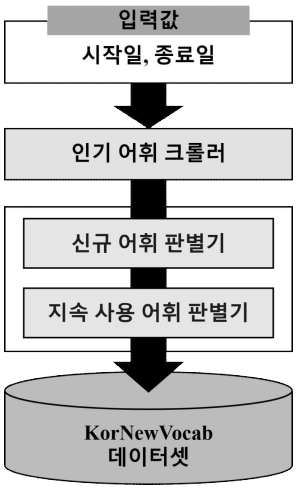



그림 1. KorNewVocab 데이터 구축 흐름 요약

이를 위해, 본 연구는 한국어 말뭉치 KorNewVocab을 새로이 제시한다. 먼저, 신조어가 가져야 할 세부 중요 요건을 1)신규 어휘, 2)인기 어휘, 3)지속 사용 어휘로 정의하고 이 조건을 만족하는 신조어 말뭉치를 2019년 1월부터 2019년 8월까지의 뉴스 기사를 중심으로 신조어 412개와 이를 포함하고 있는 4,532문장으로 구성된 신조어 말뭉치를 구축하였다.

그림 1은 KorNewVocab 데이터셋을 구축하는 과정을 간략히 표현한 흐름도이다. 또한, 본 말뭉치의 구축에 활용된 반자동 신규 어휘 검출 및 중요도 측정 방법을 소개한다.

* 교신 저자(Corresponding Author)

표 1. KorNewVocab 수집 과정

| 급상승 검색어 | 신규 어휘 | 뉴스 등장 어휘 | 검색어 트렌드 | KorNewVocab 신규어휘 |
|---|---|---|---|---|
| 코로나 바이러스 | **코로나 바이러스** | **코로나 바이러스** | **코로나 바이러스** | **코로나 바이러스** |
| 톰 행크스 | 톰 행크스 | **톰 행크스** | 톰 행크스 | 톰 행크스 |
| 미스터트롯 | **미스터트롯** | **미스터트롯** | **미스터트롯** | **미스터트롯** |
| 마스크 5부제 | **마스크 5부제** | **마스크 5부제** | **마스크 5부제** | **마스크 5부제** |
| 인계동 벤틀리 | **인계동 벤틀리** | **인계동 벤틀리** | 인계동 벤틀리 | 인계동 벤틀리 |
| 챔스 | 챔스 | **챔스** | **챔스** | 챔스 |

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로서 해외에서 진행된 신조어 관련 연구들과 국내에서 진행된 신조어 관련 연구에 대해 기술 및 신조어 데이터의 필요성에 대해 설명한다. 3장에서는 KorNewVocab 데이터셋을 구축할 때 사용한 3가지 기준에 대해 기술한다. 4장에서는 데이터 수집을 위해 사용한 프레임워크에 대해 설명하며, 5장에서는 위 과정으로 수집된 KorNewVocab을 기술 및 실험을 통해 데이터셋을 분석한다. 마지막으로 6장에서 결론을 도출하고 향후 과제를 기술한다.

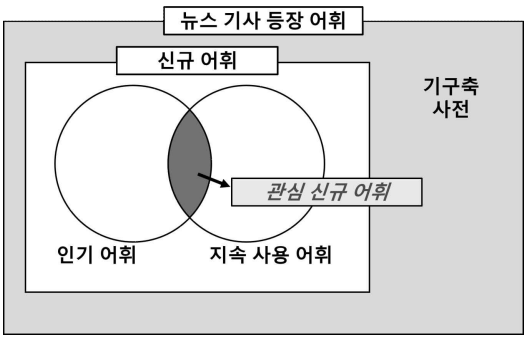



그림 2. KorNewVocab다이어그램

## 2. 관련 연구

최근 한국어에서, 많은 신조어가 만들어지고 있으며 이들이 실제 언어생활에서 차지하는 비중도 높아지고 있다. 특히 일상 구어를 넘어서 언론, 방송, 서적 등 공식적인 언어 매체에서의 빈도도 점차 증가하는 추세이다[4]. 이에 자연어 처리 연구에 활용되는 데이터셋에 신조어들이 많이 포함되며, 데이터셋에 존재하지 않는 신조어로 인한 성능 저하를 피하고자 신조어와 관련된 연구가 국내외로 활발히 진행되고 있다.

해외의 경우, 신조어를 다루기 위해 비슷한 의미의 단어들을 모아둔 데이터셋을 구축하는 연구[5]와 문장의 의미와 형태학적 특징들을 활용해 신조어의 임베딩을 진행하는 연구[6] 등 신조어와 관련된 활발히 이루어지고 있다.

국내의 경우 국립 국어원에서 배포한 한국어 학습용 어휘 목록[7]과 K-ICT 한글 형태소 사전 데이터셋[8]이 있지만, 이는 각각 2007년, 2016년 이후 최신화가 되지 않아 신조어에 관한 자료가 부족한 상황이다. 이러한 문제를 해결하기 위해 신조어의 품사를 자동으로 추정할 수 있는 형태소 분석기를 개발하는 연구[9]와 어절 분리 문제와 신조어가 포함된 문제에 대응할 수 있는 형태소 분석기 모델을 제안하는 연구[10] 등이 진행되었다.

본 연구에서는 이런 연구들의 근본적인 원인인 신조어 데이터셋의 부족을 해결하기 위한 데이터셋을 구축하기 위한 방법에 대해 논의한다.

## 3. 신조어 기준 정의

그림 2는 KorNewVocab 데이터셋에 포함된 신규 어휘를 표현한다. 이 어휘는 세 가지의 조건을 만족한 신규 어휘이며 KorNewVocab 데이터셋의 조건인 1)신규 어휘 2) 인기 어휘 3)지속 사용 어휘를 만족한다.

표 1은 전반적인 KorNewVocab 데이터셋의 수집 과정을 보여준다. 밑줄 표시된 단어들이 각 조건을 만족한 단어들이다. 정해진 기간에 사람들이 많이 검색한 검색어들을 정렬한 다음, 기존 사전에 존재하는 어휘를 제외한다. 이 과정을 통해 신규 어휘로 범위를 좁힐 수 있으며, 해당 어휘가 뉴스기사 데이터 내에서 사용되었는지 확인한다. 이렇게 추려진 후보군에서 사용 지속도와 사용 편중도 수치를 활용하여 사용자의 어휘 지속 사용 수치를 측정하여 중요도를 계산하고, 중요도가 높은 어휘들을 KorNewVocab 데이터셋의 어휘로 수집하였다.

각 기준에 대한 자세한 설명은 3.1과 3.2, 3.3에 기재하도록 한다.

### 3.1 신규 어휘

KorNewVocab 데이터셋 신규 어휘의 첫 번째 기준인 신규성이다. 신규성을 판단하기 위하여, 국립 국어원의 한국어 학습용 어휘 목록(2007)[7], K-ICT의 한글 형태소 사전(2016)[8] 및 2019년 이전 뉴스 데이터에서 등장한 어휘들을 수집해둔 사전에서 존재 여부를 확인하였다. 기구축 사전에 존재하지 않을 경우, 신조어 후보군으로 분류하여 다른 두 가지 기준을 만족하는지 확인하였다.

### 3.2 인기 어휘

KorNewVocab 데이터셋의 어휘는 사람들이 흥미를 가지며, 관심이 많은 단어를 수집하도록 하였다. [11]에서 언급한 방법과 같이 검색 포털 사이트가 제공하는 실시간 인기 검색어가 대중의 관심사를 나타내고 있는 요소로 가치가 있다고 가정하여 본 연구에서는 네이버 실시간 검색어 상위 20개의 인기 검색어를 인기 어휘 판단

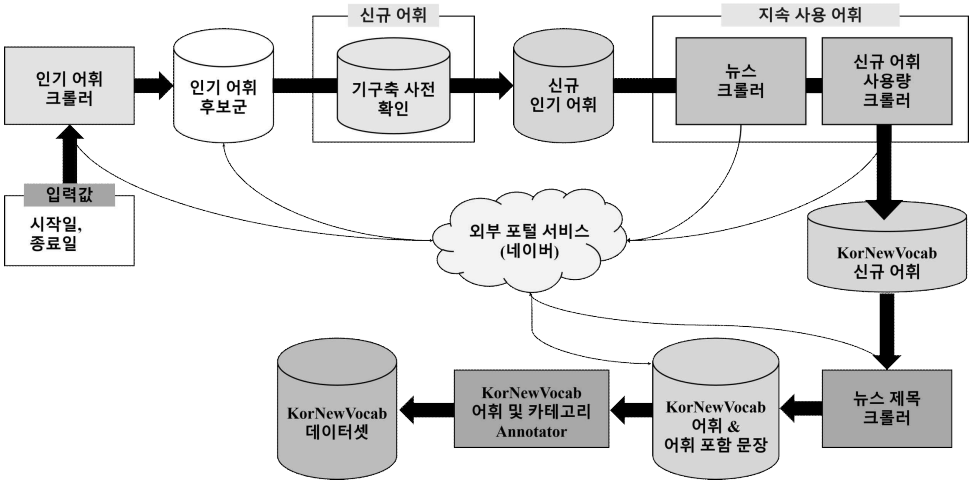



그림 3. KorNewVocab 데이터셋 구축 과정

표 2. 2020.09.03. 오전 8시 네이버 실시간 급상승 검색어 상위 10개

| Rank | 실시간 급상승 검색어 |
|---|---|
| 1 | 하이선 |
| 2 | 태풍 10호 하이선 |
| 3 | 정전 |
| 4 | 이혜성 |
| 5 | 한전 |
| 6 | 태풍 마이삭 서울 |
| 7 | 네이마르 |
| 8 | 에이전트h |
| 9 | 강릉 |
| 10 | 태풍 피해 |

기준으로 활용하였다. 인기 어휘 판단 기준으로 사용한, 네이버 플랫폼은 국내 검색어의 70% 이상을 수용하여 활용하기에 적합하다고 판단하였다.

표 2는 2020년 9월 3일 오전 8시의 급상승 검색어를 보여준다. 이를 통해, 당일 태풍 하이선이 사람들이 가장 많이 검색 및 사용한 단어이며 당시 인기가 많았던 단어라는 것을 알 수 있다.

### 3.3 지속 사용 어휘

본 연구에서는 신규 어휘가 일시적으로 사용되는 것이 아닌, 지속적으로 고르게 사용되는 것이 중요하다고 판단하였다. 이를 확인하기 위하여, [12]에서 제안한 세 가지 척도 중 ***사용지속도***와 ***사용편중도***를 활용하여 사용도를 확인하였다. 사용 지속도와 사용 편중도에서 활용되는 사용량은 해당 단어가 등장한 뉴스 기사 건수로 정의한다. 각 지표는 아래와 같이 계산된다.

**사용지속도 :** 꾸준히 사용되는 단어를 구분 짓기 위한 척도이다. 단어의 사용 지속도는 (해당 단어의 사용량이 1 이상인 기간 수)/(총 기간 수)로 정의된다.

**사용편중도 :** 특정 시기에 매우 많은 양이 사용되지 않다가 특정 시기에만 매우 많이 사용되는지 측정할 수 있는 척도이다. 편중도는 (해당 단어의 최대 사용량)/(해당 단어의 평균 사용량)으로 계산한다.

표 3. 신조어 카테고리

| 카테고리 명 | 설명 | 예제 |
|---|---|---|
| PS | 사람 이름 | 홍길동, 아이유, 짱구 |
| OG | 기관/단체/업체 | 공정거래위원회,한국은행 |
| EV | 특정 사건/사고 명칭 | 새마을운동, 88서울올림픽 |
| AF | 사람에 의해 창조된 대상물 | 경복궁, 피아노, 동의보감 |
| CV | 문명/문화 관련 명칭 | 안데스문명, 유대인, 쓰레기종량제 |
| TM | 앞서 정의된 개체명 이외의 개체명 | 아열대 기후, 리눅스, 인스타그램 |

### 3.4 중요도

본 연구에서는 [12]에서 정의한 사용 지속도와 사용 편중도를 활용하여 중요도라는 새로운 척도를 제안한다.

**중요도**: 단어가 지속적으로 나타나며, 평균 사용량이 높은 단어가 중요하다고 판단하여, 중요도는 평균 사용량 * 1/(사용 편중도)로 정의한다. 평균 사용량은, 단어가 등장한 이후, 해당 단어가 등장한 뉴스 기사 건수의 평균값이다.

### 3.5 신조어 개체명 카테고리

앞서 설명한 과정을 거쳐 수집한 KorNewVocab 데이터셋의 신규 어휘를 분류하는 카테고리에 대해 설명한다. 카테고리는 한국정보통신기술협회에서 배포한 개체명 태그 세트[13] 및 태깅 말뭉치[6]와 창원대학교, 한국 해양 대학교 자연언어처리 연구실에서 배포한 개체명 분류 카테고리를 참고하여 작성하였다.

신조어 카테고리는 표 3과 같이 PS, OG, EV, AF, CV, TM 여섯 가지 카테고리로 구성되어 있다.

표 4. KorNewVocab 분석 결과

| Figure | Count |
|---|---|
| 어휘 당 문장 수 | 11 |
| char 단위 문장 평균 길이 | 32 |
| word 단위 문장 평균 길이 | 7 |
| 코모란 형태소 분석기 평균 분할 개수 | 15 |
| 한나눔 형태소 분석기 평균 분할 개수 | 11 |
| 트위터 형태소 분석기 평균 분할 개수 | 14 |
| 꼬꼬마 형태소 분석기 평균 분할 개수 | 17 |
| 총 문장 수 | 4,532 |

## 4. 데이터 수집 프레임워크

그림 3은 KorNewVocab을 구축하는 전체적인 과정을 나타낸다.

### 4.1 인기 어휘

3장에서 기술하였던 것과 같이 사람들의 사용량이 높은 단어를 수집하기 위하여 네이버 실시간 급상승 검색어를 수집하여 인기 어휘 조건을 만족하는 후보군들을 수집하였다. 3시간 단위로 수집을 진행하여 2019년 1월부터 8월까지 검색량이 많은 단어를 수집하였다.

### 4.2 신규 어휘

기구축 사전에 등록되어있던 어휘들은 제거하여 KorNewVocab 데이터셋의 신규 어휘 후보군을 정렬하였다. 기구축 사전은 국립 국어원의 한국어 학습용 어휘 목록(2011), K-ICT의 한글 형태소 사전(2017) 및 2019년 이전 뉴스 데이터에서 등장한 어휘들을 수집해둔 어휘 사전이다.

### 4.3 지속 사용 어휘

지속적인 사용도 평가 척도인 사용 지속도와 사용 편중도에 사용되는 사용량을 수집한다. 사용량은 [12]에서 정의한 해당 단어가 등장한 뉴스 기사 건수를 활용하였다. KorNewVocab는 2달 이상 사용된 어휘를 선택하기 위하여 사용 지속도는 0.25 이상, 사용 편중도는 8 이하인 어휘를 대상으로 수집하였다.

### 4.4 신규 어휘 포함 문장 수집

세 가지 조건을 모두 만족한 신규 어휘를 포함하고 있는 문장을 수집하기 위하여 프레임워크에서 같은 기간 내 단어를 포함하고 있는 뉴스 기사 제목을 수집하도록 한다. 동음이의어인 경우는 제외하여 수집하였다.

### 4.5 문장 내 신규 어휘 태깅 및 카테고리 분류

4.4 단계를 거쳐 수집된 문장에 포함된 신규 어휘에 개체명 말뭉치에서 개체명을 인식하기 위한 가장 보편적인 방법의 하나인 BIO 태깅을 진행하였다.

표 5. KorNewVocab 데이터셋의 신규 어휘들을 한국어 형태소 분석기로 추출한 결과

| 한국어 형태소 분석기 | 신규 어휘 추출 횟수 | 성능 |
|---|---|---|
| 코모란 | 1,003 | 22.13% |
| 한나눔 | 1,420 | 31.33% |
| 트위터 | 741 | 16.35% |
| 꼬꼬마 | 107 | 2.36% |
| 신규 어휘 총합 | 4,532 | |

표 6. KorNewVocab 데이터셋의 신규 어휘들을 한국어 형태소 분석기로 추출하였을 때 타입 별 결과

| 한국어 형태소 분석기 | 명사형 어휘 추출 성능 | 명사구형 어휘 추출 성능 |
|---|---|---|
| 코모란 | 23.75% | 18.73% |
| 한나눔 | 46.27% | 0% |
| 트위터 | 24.14% | 0% |
| 꼬꼬마 | 3.48% | 0% |

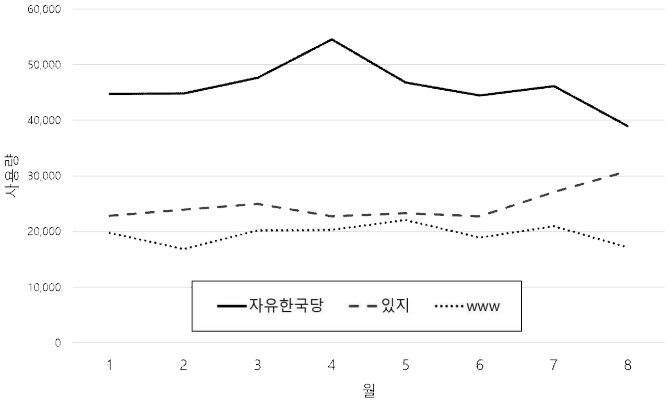
## 5. KorNewVocab 데이터셋 실험 및 분석

4장에서 기술한 프레임워크를 사용하여 총 412개의 신규 어휘들을 수집하였다. 단어별로 11문장 씩 수집하였으며, 이렇게 수집된 문장은 4,532문장이다. 412개의 단어들 중 279개는 명사형, 나머지 133개는 명사구형 데이터로 구성되었다. 표 4는 KorNewVocab 데이터셋을 전체적으로 분석한 결과이다.

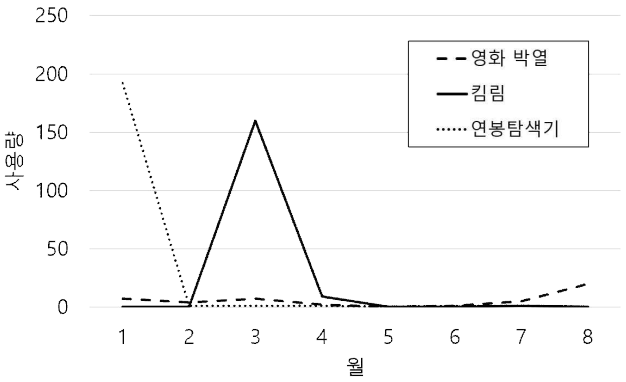
코모란, 한나눔, 트위터 그리고 꼬꼬마는 대표적인 한국어 형태소 분석기이다. KorNewVocab의 신규 어휘들의 성능을 분석하기 위해 다음과 같은 형태소 분석기들을 사용하였다.

표 5는 본 연구에서 제작한 KorNewVocab 데이터셋을 통하여 과거 데이터로 개발된 형태소 분석기들이 신조어를 분석하기에 적합하지 않음을 증명하기 위해 실험을 진행한 결과이다.

표 6는 신규 어휘의 유형에 따라 한국어 형태소 분석기의 성능을 분석한 것이다. 코모란 형태소 분석기를 제외한 형태소 분석기들은 명사구형 단어에 취약함을 확인할 수 있었다.




(a) 상위 중요도 신규 어휘 월별 사용량




(b) 하위 중요도 신규 어휘 월별 사용량

그림 4. 신규 어휘 월별 사용량

그림 4(a)는 KorNewVocab 데이터셋 내 중요도 상위 3위 안의 단어들의 월별 사용량을 나타낸 도표이다. 중요도가 높은 단어들은 평균 사용량이 많으며, 월별로 고르게 분포되어 사용 편중도가 낮은 것을 알 수 있다.

또한, 그림 4(b)를 통해 사용도가 높더라도 고르게 사용되지 않거나 사용량이 적은 신규 어휘의 중요도가 낮은 것을 확인할 수 있었다.

표 7은 KorNewVocab 데이터셋의 일부이다. 신조어와 신조어를 포함하는 예시 문장, 그리고 신조어와 카테고리가 태깅되어 있는 BIO 태깅 결과이다. 이 때, BIO 태깅이란 개체명의 시작을 “B”로, 개체명의 이어지는 부분을 “I”로, 개체명이 아닌 경우 “O”로 태깅하는 방식이다. 그리고 카테고리 명을 개체명의 시작인 “B”와 함께 표기한다.

이렇게 구성된 KorNewVocab 데이터셋은 연구 목적으로 배포할 계획이다.

표 7. KorNewVocab 데이터셋

| | |
|---|---|
| OOV | 노노재팬 |
| 문장 | 뱃길도 노노재팬・・・부산～대마도 여객선 잇단 운항 중단 |
| 태깅 결과 | 뱃길도 <EV>노노재팬</EV>・・・부산～대마도 여객선 잇단 운항 중단 |

(a) 노노재팬

| | |
|---|---|
| OOV | 당근마켓 |
| 문장 | 당근마켓, 네이버 라인 출시 앱 표절 의혹 제기 |
| 태깅 결과 | <TM>당근마켓</TM>, 네이버 라인 출시 앱 표절 의혹 제기 |

(b) 당근마켓

| | |
|---|---|
| OOV | 라라랜드 |
| 문장 | 서울광장서 25일 '라라랜드' 상영…26・27일 '빗물콘서트' |
| 태깅 결과 | 서울광장서 25일 '<AF>라라랜드</AF>' 상영…26・27일 '빗물콘서트' |

(c) 라라랜드

## 6. 결론 및 향후 과제

KorNewVocab는 세 가지 신규 어휘 조건인 신규 어휘, 인기 어휘, 지속 사용 어휘를 만족하는 단어들을 수집한 데이터셋이다. 총 412개의 신규 어휘와 이를 포함하는 4,532개의 문장으로 구성되었다. 또한, 신규 어휘의 중요도라는 새로운 지표를 제시하여 신규 어휘의 중요성을 판단할 수 있는 척도로 활용할 수 있도록 하였다.

본 연구는 이러한 조건들을 사용하여 데이터를 수집하는 반자동 프레임워크와 수집된 단어들을 실험 및 분석하였으며, 이를 통해 기존 형태소 분석기들이 신규 어휘 분석에 적합하지 않음을 증명하고 형태소 분석기들의 개선을 위한 데이터셋을 구축하였다.

제안한 프레임워크와 세 가지 기준, 중요도 척도를 활용하여 제작된 높은 완성도의 KorNewVocab 데이터셋은 NLP 연구에서 신규어휘로 인한 성능 저하를 막기 위한 데이터셋으로 응용될 수 있을 것이다.

## 감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-01441)

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00004, 준지도학습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발)

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2019R1F1A1060601)

## 참고문헌


[1] 표시연. 영어 형태소가 합성된 신조어 생성의 형태론적 유형에 대한 고찰. 언어, 42(1), 97-120. (2017).

[2] He, Keqing, Yuanmeng Yan, and X. U. Weiran. "Learning to Tag OOV Tokens by Integrating Contextual Representation and Background Knowledge." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

[3] Kolachina, Prasanth, Martin Riedl, and Chris Biemann. "Replacing OOV words for dependency parsing with distributional semantics." Proceedings of the 21st Nordic Conference on Computational Linguistics. 2017.

[4] 김보현. "신조어를 활용한 한국어 단어 형성법 교육 내용 연구." 국제한국어교육학회 춘계학술발표논문집 2020 (2020): 204-220.

[5] Lazaridou, Angeliki, Marco Marelli, and Marco Baroni. "Multimodal word meaning induction from minimal exposure to natural text." Cognitive science 41 (2017): 677-705.

[6] Hu, Ziniu, et al. "Few-Shot Representation Learning for Out-Of-Vocabulary Words." arXiv preprint arXiv:1907.00505 (2019).

[7] 김흥규, 강범모, 홍정하. 21세기 세종계획 현대국어 기초말뭉치: 성과와 전망. 한국정보과학회 언어공학 연구회 학술발표 논문집, 311-316. (2007).

[8] NIA(National Information Society Agency), Jeon H. NIADic: NIA(National Information Society Agency) Korean Dictionaries. R package version 0.0.1. https://github.com/haven-jeon/NIADic. Accessed 8 Apr 2019, (2016).

[9] 박한신, 임소라, 권용진. AI 기반의 신조어 자동 태깅 구현을 통한 HMM 형태소 분석기의 고성능화. 한국통신학회 학술대회논문집, 834-835.(2018).

[10] 최병서, 이익훈, 이상구. 신조어 및 띄어쓰기 오류에 강인한 시퀀스-투-시퀀스 기반 한국어 형태소 분석기. 정보과학회논문지, 47(1), 70-77. (2020).

[11] 김시원, 손재기, 안재훈. 실시간 고속 빅데이터 처리 플랫폼 기반 인기 검색어 뉴스 크롤링을 활용한 사회적 이슈 추출 및 검색어 모델링. 정보 및 제어 논문집, 311-312. (2018).

[12] 한경수. 뉴스 기사에서 지속도와 편중도 기반의 신조어 사용 특징 분석. 한국디지털콘텐츠학회 논문지, 20(1), 51-58.(2019).

[13] TTA, “Tag Set and Tagged Corpus for Named Entity Recognition”, TTAK.KO-10.0852, (2015)