

형태소 분석기 활용 방안

1. 형태소 분석의 목적과 활용

형태소 분석의 목적

- 형태소 분석은 통사 분석과 의미 분석의 전단계로서 기계 번역이나 자연 언어 인터페이스, 정보 검색 등 모든 자연언어 관련 분야에 필수적이다. 특히 한국어, 일러처럼 문법형태소의 기능에 의해 단어의 통사적, 의미적 역할이 결정되는 교착어에서는 형태소 분석이 통사 분석과 의미 분석에 미치는 영향이 크기 때문에 한국어의 분석에 있어서 형태소 분석이 아주 중요하다고 할 수 있다.
- 전산학적인 관점에서 형태소 해석은 어절에 대한 사전 정보를 적재하기 위하여 어절을 구성하는 사전 표제어 혹은 형태소를 파악하여 적절한 사전 정보를 적재하는 것이다.
- 형태소 분석은 자연언어 처리에 필요한 정보를 추출하는 기저 단계로 처리 시스템의 초석으로써 형태소 분석의 기능과 효율성은 전체 시스템에 바로 직결되는 중요한 요소인 것이다

기존 형태소 분석기의 활용

- 한국어 NLP에서 형태소 분석기를 사용한다는 것은 단어 토큰화가 아니라 정확히는 형태소(morpheme) 단위로 형태소 토큰화(morpheme tokenization)를 수행하게 됨을 뜻합니다.

```
# 토큰화의 예시
from konlpy.tag import Okt
okt=Okt()

print(okt.morphs("열심히 코딩한 당신, 연휴에는 여행을 가봐요"))
>>> ['열심히', '코딩', '한', '당신', ',', '연휴', '에는', '여행', '을', '가봐요']

print(okt.pos("열심히 코딩한 당신, 연휴에는 여행을 가봐요"))
>>> [('열심히', 'Adverb'), ('코딩', 'Noun'), ('한', 'Josa'), ('당신', 'Noun'), (',', 'Punctuation'), ('연휴', 'Noun'), ('에는', 'Josa'), ('여행', 'Noun'), ('을', 'Josa'), ('가봐요', 'Verb')]

print(okt.nouns("열심히 코딩한 당신, 연휴에는 여행을 가봐요"))
>>> ['코딩', '당신', '연휴', '여행']
```

• 활용 예시

1. 기계번역
2. 검색엔진
3. 그 밖에 자연언어 처리에 필요한 정보를 추출하는 기저 단계로 활용할 경우 모두

2. Three cow tagger의 강점

- 신조어 추가 + 이모티콘 분류 + 해시태그 처리
- 인스타에 특화되어있음.

* okt 참고

```
print(oks.pos("머해 진짜! 완전 짱응나게해 아니 하.. 화내지말자 ㄹㅇ 존맛탱 오오티디😄😄 #해쉬태그"))
>>> [('머', 'Noun'), ('해', 'Verb'), ('진짜', 'Noun'), ('!', 'Punctuation'),
('완전', 'Noun'), ('짱응', 'Noun'), ('나게해', 'Verb'), ('아니', 'Adjective'),
('하', 'Exclamation'), ('..', 'Punctuation'), ('화', 'Noun'), ('내지말자', 'Verb'),
('ㄹㅇ', 'KoreanParticle'), ('존맛', 'Noun'), ('탱', 'Adverb'),
('오오티디', 'Noun'), ('티디', 'Noun'), ('😄😄', 'Foreign'), ('#해쉬태그', 'Hashtag')]
```

띄어쓰기만 잘해주면 거의 정확히 분류해낸다.

1. [ㄹㅇ, 존맛탱, 오오티디] 와 같은 신조어를 잘 분류하지 못한다. → 사전추가
2. [😄😄] 이모지의 경우 따로 분류하지 않는다. → 이모지 분류 품사 추가
ex) ('😄_smile face', 'Emoji')
3. [#해쉬태그] #뒤의 문자를 전부 Hashtag로 분류한다.

3. Three cow tagger 활용 방안

~~사실상 Okt 업그레이드 버전이므로 활용 방안 역시 같다...(SNS 특화)~~

• 통계 관련

1. 검색 해시태그 트렌드 지수 → 최근 자주 사용된 해시태그 분석
2. 인기 포스트 분석 통계 → 조회수가 높은 포스트 텍스트 분석
3. 시간대별 게시물 관련 통계 → 시간대별로 자주 쓰인 텍스트 분석
4. 연관/추천 해시태그 클라우드 → 같이 사용될 가능성이 높은 해시태그 제안

• 유저 프로파일링

1. best 인게이지먼트 포스트 → 소비자와 관련성이 높은 포스트의 형태 제안
2. 콘텐츠 추천 시스템 → AI를 활용한 특정 유저의 관심 콘텐츠 추천(영화, 책, 쇼핑)
→ 비슷한 사용자의 경험을 토대로 예측
3. 소개팅, 모임 주선 시스템
4. 맞춤형 인공지능 챗봇 → 상대방의 관심사, 연령 등을 토대로 대화하는 맞춤형 챗봇

• 그 외 활용 방안