

*** 사용 방법 ***

1. 실행 명령어

1) 기본 명령어

./moHANA

예)1 학교는	학교_{ncn} + 는_{j}
학생및선생	학생_{ncn} + 및_{ad} + 선생_{ncn}
안적는	안적는_{unk}
엄마와아이	엄마_{ncn} + 와아이_{unk}

2) 옵션 명령어

./moHANA -2

: 흔히 잘못 쓰는 오류를 처리한 형태소 분석

예)2 학교는	학교_{ncn} + 는_{j}
학생및선생	학생_{ncn} + 및_{ad} + 선생_{ncn}
안적는	안_{ad} + 적_{pv} + 는_{ef}
엄마와아이	엄마_{ncn} + 와_{j} + 아이_{ncn}

2. 출력형태

moHANA의 모든 형태소들은 5차원 언어적 자질들을 갖고 있으나, i moHANA 베타 i 버전에서의 형태소 분석된 어절의 언어 정보 출력 형태는 1차원적으로 i 어류 태그(class of tag) i 만을 출력하고 있다.

*** 분석 기준 ***

1. **moHANA**는 원칙적으로 띄어쓰기 및 한국어 문법에 맞게 쓰여진 어절들을 형태소 분석의 대상으로 삼는다.
2. 가능한 한국어 문법에 적절한 형태소 분석을 하는 것을 원칙으로 한다.
3. 그러나, 말뭉치에서 흔히 잘못 쓰고 있는 어절들도 형태소 분석할 수 있는 장치를 마련해 놓았다.

- 예)1 잘먹는 잘_{ad} + 먹_{pv} + 는_{ef}
 안예빠지는 안_{ad} + 예빠_{pa} + 어_{ef} + 지_{aux} + 는_{ef}
- 못자르는 못_{ad} + 자르_{pv} + 는_{ef}
 안미끄러운 안_{ad} + 미끄럽_{pa} + 은_{ef}
- 예)2 편지에대한 편지_{ncp} + 에_{j} + 대하_{pv} + ㄴ_{ef}
 청주에가는방법 청주_{ncn} + 에_{j} + 가_{pv} + 는_{ef} + 방법_{ncn}
- 생존하기위한방법 생존_{ncp} + 하_{vfix} + 기_{ef} + 위하_{pv} + ㄴ_{ef} + 방법_{ncn}
- 태풍으로인한피해 태풍_{ncn} + 으로_{j} + 인하_{pv} + ㄴ_{ef} + 피해_{ncp}
- 예)3 먹지못하는 먹_{pv} + 지_{ef} + 못하_{aux} + 는_{ef}
- 예)4 책상과의자 책상_{ncn} + 과_{j} + 의자_{ncn}
- 이효리의남자친구 이효리_{nq_per} + 의_{j} + 남자_{ncn} + 친구_{ncn}

4. 특히, 법률 용어에 대한 형태소 분석이 용이하다.

- 예)4 지가공시및토지등의평가에대한법률
 지가공시_{ncp} + 및_{ad} + 토지_{ncn} + 등_{nfix} + 의_{j} + 평가_{ncp} + 에_{j} + 대하_{pv} + ㄴ_{ef} + 법률_{ncn}
- 성매매알선등행위의처벌에대한법률
 성매매_{ncn} + 알선_{ncp} + 등_{nfix} + 행위_{ncn} + 의_{j} + 처벌_{ncp} + 에_{j} + 대하_{pv} + ㄴ_{ef} + 법률_{ncn}

5. 접미사와 접두사를 세분화하여 형태소 분석에 활용하고 있다.

- 예)5 서귀포시를 서귀포_{nq_loc} + 시_{nfix} + 를_{j}
- 서귀포역을 서귀포_{nq_loc} + 역_{nfix} + 을_{j}

서귀포점을	서귀포_{nq_loc} + 점_{nfix} + 을_{j}
예)6 (주)마이크로소프트	(주)_{pref} + 마이크로_{nq_gro} + 소프트_{ncp}
(주)포항제철	(주)_{pref} + 포항_{nq_gro} + 제철 _{ncp}
(주)컴퓨터	(_{ascii} + 주_{nc_one} +)_{ascii} + 컴 퓨터_{ncn}

(알림)

1. 현재 릴리즈하려고 하는 **jmoHANA** 베타j는 비영리적인 목적으로 사용하는 분들을 위한 것입니다.
따라서 **jmoHANA** 베타j는 영리적 목적으로 사용할 수 없습니다.
2. **moHANA**는 현재 특허 출원 중입니다.
(출원번호 : 10-2007-0024439)
3. 현재 시험되고 있는 **moHANA**는 베타 버전입니다.
4. 혹시 형태소 분석의 정확률과 형태소 분석의 속도에 대한 문의가 있으면 연락 주십시오.
(현재 시험되고 있는 **jmoHANA** 베타j의 기본식 사전의 어절수는 약 13만여 어절임. 40만절 어절까지 늘릴 예정임.)
5. **moHANA** 베타 버전을 **binary** 버전으로 배포하고 있으니, 이를 사용하고 싶은 분은 아래의 연락처로 연락해 주십시오..
5. **moHANA**를 사용하여 형태소 분석을 하다가 불편한 점이나 잘못된 사항을 알려 주시면 **moHANA**를 발전시키는 데에 많은 도움이 되니, 언제라도 연락을 해 주시기 바랍니다..
6. 연락처: shsuh@wordwords.co.kr