

Comparison of Morpheme Analysis

Mecab vs Khaiii

Details

- Morpheme unit improves word embedding quality in Korean.
- Object : Korean National Anthem

Condition

- Eliminate whole spacing words

• 애국가

1. 동해물과 백두산이 마르고 닳도록
하느님이 보우하사 우리나라 만세
무궁화 삼천리 화려강산
대한 사람 대한으로 길이 보전하세

2. 남산 위에 저 소나무 철갑을 두른 듯
바람서리 불변함은 우리 기상 일세
무궁화 삼천리 화려강산
대한 사람 대한으로 길이 보전하세

3. 가을 하늘 공활한데 높고 구름 없이
밝은 달은 우리 가슴 일편단심 일세
무궁화 삼천리 화려강산
대한 사람 대한으로 길이 보전하세

4. 이 기상과 이 맘으로 충성을 다하여
괴로우나 즐거우나 나라 사랑하세
무궁화 삼천리 화려강산
대한 사람 대한으로 길이 보전하세

• Mecab

1. 동해 물 과 백두산 이 마르 고 닳 도록
하느님 이보우 하사 우리나라 만세
무궁화 삼천리 화려 강산
대한 사람 대한 으로 길이 보전 하 세

2. 남산 위 에 저 소나무 철갑 을 두른 듯
바람서리 불변 함 은 우리 기상 일세
무궁화 삼천리 화려 강산
대한 사람 대한 으로 길이 보전 하 세

3. 가을 하늘 공활 한 데 높 고 구름 없이
밝 은 달 은 우리 가슴 일편단심 일세
무궁화 삼천리 화려 강산
대한 사람 대한 으로 길이 보전 하 세

4. 이 기상 과 이맘 으로 충성 을 다 하 여
괴로우 나 즐거우 나 나라 사랑 하 세
무궁화 삼천리 화려 강산
대한 사람 대한 으로 길이 보전 하 세

• Khaiii

1. 동해물 과 백두산 이 마르 고 닳 도록
하느님 이 보 우 하 사우리나라만세
무궁화삼천 리화려강산
대 하 ㄴ 사람대한 으로 길이보전하세

2. 남산 위 에 저 소나무철갑 을 두르 ㄴ 듯
바 람 서 리 불변함 은 우리 기상일세
무궁화삼천 리화려강산
대 하 ㄴ 사람대한 으로 길이보전하세

3. 가을하늘 공활 하 ㄴ 데 높 고 구름없이
밝 은 달 은 우리 가슴 이 ㄴ 편단심일세
무궁화삼천 리화려강산
대 하 ㄴ 사람대한 으로 길이보전하세

4. 이 기상 과 이맘 으로 충성 을 다하 여
괴로우 나 즐 거 우 나 나라사랑 하 세
무궁화삼천 리화려강산
대 하 ㄴ 사람대한 으로 길이보전하세

Results

- Mecab beats Khaiii.
- But not flawless (somewhat incomplete)