

Lyrics Analyzer™

MATLAB 잘하쥬 3조 / 조교 조예진

공은광 송보미 김정욱 최윤서 신유경

Contents

- **What is Lyrics Analyzer**
- **Algorithms**
- **Demo**

What is Lyrics Analyzer

DB-based Lyrics Analyzer

Web-based Lyrics Analyzer

(+) Lyrics Estimator

DB-based Lyrics Analyzer

- **lyricsDB**
 - 2,275 songs
 - 154 artists in Korea
 - 1933 to present
 - Year, Title, Artist, Lyrics

Lyrics Analyzer

DB-based ver.

POS tag

☒ OFF
☐ ON

Artist & Year

☒ Artist

10cm

2NE1
2PM
AOA
BIGBANG
DJ DOC
EXID
EXO
H.O.T
S.E.S.
YB
Zion.T
f(x)
miss A
강산에
강수지
거미
거북이
걸스데이
공일오비

☐ Year

~

Start

List of Songs

by Popularity

	Year	Title	Artist	Lyrics
1				
2				
3				
4				

Basic Statistics

	Index	Result
1		
2		
3		
4		

Frequency Analysis

☒ Unigram

☐ Bigram

☒ Noun
☐ Verb
☐ Adj

Analyze

	Words	Count
1		
2		
3		
4		

Create Wordcloud

Lyrics Analyzer

DB-based ver.

POS tag ☒ OFF
☐ ON

Artist & Year

☒ Artist

10cm

2NE1
2PM
AOA
BIGBANG
DJ DOC
EXID
EXO
H.O.T
S.E.S.
YB
Zion.T
f(x)
miss A
강산에
강수지
거미
거북이
걸스데이
공일오비

☐ Year

~

Start

List of Songs

by Popularity

	Year	Title	Artist	Lyrics
1	2014	쓰담쓰담	10cm	알고 있지 오늘따라 그대가 ...
2	2015	10월의 날씨	10cm	오늘의 날씨는 그리 맑지 않...
3	2014	스토커	10cm	나도 알아 나의 문제가 무엇...
4	2014	그리워라	10cm	가끔씩 자꾸 니 생각이 나서...
5	2011	사랑은 은하...	10cm	사랑은 은하수 다방 문 앞에...
6	2012	애상	10cm	말리지마 내 이런 사랑을 너...
7	2015	비밀연애	10cm	잠깐만요 살다보면 작은 비...
8	2011	그게 아니고	10cm	그게 아니고 어두운 밤 골목...
9	2011	안아줘요	10cm	비오는 날 어느 그날 밤 같...
10	2010	Good Night	10cm	오늘 무슨 일이 생길 것만 ...
11	2014	짝사랑	10cm	그대의 표정이 너무 차가와...
12	2012	Fine Thank...	10cm	너의 애깁 들었어 너는 벌써...

Basic Statistics

	Index	Result
1	Words per song	119.13
2	Syllables per song	326.53
3	Repeated words per song	13.54
4	Total Repetitions	26.59
5	Foreign word ratio(%)	4.53

Frequency Analysis

☒ Unigram

☐ Bigram

☒ Noun
☐ Verb
☐ Adj

Analyze

	Words	Count
1	귓가에	3
2	나서	3
3	누가	3
4	없지	3
5	가고	2
6	거리를	2
7	건지	2
8	괜히	2
9	그게	2
10	그대의	2
11	그리	2
12	근데	2
13	꿈속에	2
14	나쁜	2
15	내내	2
16	다리	2
17	다시는	2

Create Wordcloud

Lyrics Analyzer

DB-based ver.

POS tag ☐ OFF ☒ ON

Artist & Year

☒ Artist

10cm

2NE1
2PM
AOA
BIGBANG
DJ DOC
EXID
EXO
H.O.T
S.E.S.
YB
Zion.T
f(x)
miss A
강산에
강수지
거미
거북이
걸스데이
공일오비

☐ Year

~

Start

List of Songs

by Popularity

	Year	Title	Artist	Lyrics
1	2014	쓰담쓰담	10cm	알=VV,고=EC,있=VV,지=...
2	2015	10월의 날씨	10cm	오늘=NNG,의=JKG,날씨...
3	2014	스토커	10cm	나=NP,도=JX,알=VV,아=...
4	2014	그리워라	10cm	가끔=MAG,씩=XSN,자꾸...
5	2011	사랑은 은하...	10cm	사랑=NNG,은=JX,은하수...
6	2012	애상	10cm	말리=VV,지=EC,말=VX,...
7	2015	비밀연애	10cm	잠깐=MAG,만=JX,요=JX...
8	2011	그게 아니고	10cm	그=NP,게=EC,아니=VCN...
9	2011	안아줘요	10cm	비오=NNP,는=JX,날=NN...
10	2010	Good Night	10cm	오늘=NNG,무슨=MM,일...
11	2014	짝사랑	10cm	그대=NP,의=JKG,표정=N...
12	2012	Fine Thank...	10cm	너=NP,의=JKG,얘기=NN...
13	2010	새벽 4시	10cm	가라기=VV,기=EC,참으...

Basic Statistics

	Index	Result
1	Words per song	119.13
2	Syllables per song	326.53
3	Repeated words per song	13.54
4	Total Repetitions	26.59
5	Foreign word ratio(%)	4.53

Frequency Analysis

☒ Unigram

☐ Bigram

☐ Noun

☐ Verb

☒ Adj

Analyze

	Words	Count
1	싫다	4
2	어지럽다	4
3	그립다	3
4	아름답다	3
5	나쁘다	2
6	두렵다	2
7	뜨겁다	2
8	부끄럽다	2
9	슬프다	2
10	어떻다	2
11	어렵다	2
12	예쁘다	2
13	작다	2
14	차갑다	2
15	괜찮다	1
16	귀찮다	1
17	깊다	1

Create Wordcloud

Web-based Lyrics Analyzer

- **function**
ArtistSongs
 - [Input] artistName, nSongs
 - [Output] Year-Title-Artist-Lyrics table
 - web access to genie.co.kr
 - Open to any input a user wants

Lyrics Analyzer

Web-based ver.

Choose your artist

Artist

Number of Songs
for Analysis

Start

Specify years in interest (optional)

Year

~

Refine

Reload full data

Basic Statistics

	Index	Result
1		
2		
3		
4		

Save

Load

Add

POS

☒ OFF

☐ ON

List of Songs

by Popularity

	Year	Title	Artist	Lyrics
1				
2				
3				
4				

Frequency analysis

☐ Unigram

☒ Bigram

☒ Noun

☐ Verb

☐ Adj

Analyze

	Words	Count
1		
2		
3		
4		

Create Wordcloud

(+) Lyrics Estimator

Type in lyrics

Estimated Year :

Estimated Artist :

	1	2
1		
2		
3		
4		

Estimate!

Search DB

Update DB

Algorithms

web-based

Algorithms



(1) Input via Web



(2) Basic Statistics



(3) Pre-processing



(4) POS tag



(5) Get specific POS



(6) Stopwords removal



(7) Frequency Analysis



(8) Wordcloud

데이터 수집

기본 통계치(5가지) 계산

전처리 (노이즈 제거)

형태소 분석

특정 형태소 추출

불용어 처리

빈도 분석

워드클라우드 그리기

(1)Input via Web

ArtistSongs()

```
=function [ fullData ] = ArtistSongs(artistName, nSongs)
% This function retrieves Year-Title-Artist-Lyrics information
% of a requested number of songs of a specific artist.
% [ input  ]  artistName : name of a specific artist (char; in quotations)
%              nSongs    : number of songs to be collected (double)
% [ output ]  fullData   : n by 4 cell array of year-title-artist-lyrics
```

- ✓ Skills used
 - **webread** (cf. Get & Post)
 - **regexp**

(2) Basic Statistics

Basic Statistics		
	Index	Result
1	Words per song	81.97
2	Syllables per song	218.60
3	Repeated words per song	8.60
4	Total Repetitions	18.30
5	Foreign word ratio(%)	4.27

단어 수

음절 수

반복단어 수

반복 횟수

외국어 비율

✓ Skills used

- strcmp, regexp
- unique, categorical, countcats

(3)Pre-processing

	1	2	3	4
1	1996	3!4!	룰라	song 랄랄~ rap 그대와 우리 모두 우 기쁜 미래...
2	1995	날개 잃은 천사	룰라	** 아 그럴꺼야 나를 아꼈려고 굳이 내게 말안하고...
3	1997	비밀은 없어	룰라	우우우우우~ 우우우우우~ 우우우우우~ 우우우우우...
4	1997	연인	룰라	높은 높은 하늘을 봐봐 lover lover lover grow ...
5	1995	천상유애	룰라	rap> 저 끝에서 끝까지 목이 메이도록 말하고 말꺼...
6	1995	프로와 아마추		할수가 있어 너가 나를 ...

lyricsPrep()

	1	2		4
1	1996	3!4!	룰라	랄랄 그대와 우리 모두 우 기쁜 미래를 향해 랄랄 내가 힘...
2	1995	날개 잃은 천사	룰라	아 그럴꺼야 나를 아꼈려고 굳이 내게 말안하고 멀리 떠...
3	1997	비밀은 없어	룰라	우우우우우 우우우우우 우우우우우 우우우우우 우우 정...
4	1997	연인	룰라	높은 높은 하늘을 봐봐 lover lover lover grow 높은 ...
5	1995	천상유애	룰라	저 끝에서 끝까지 목이 메이도록 말하고 말꺼야 무척 소...
6	1995	프로와 아마추어	룰라	나는 나는 모든것을 할수가 있어 너가 나를 사랑하게 할...

✓ Skills used

- (tons of) regexprep

(4) POS tag

- No Korean POS tagger source found in Matlab!

KOMORAN 2.0 - 자바 기반의 한국어 형태소 분석기

개요

KorPOSTag()

SHINEWARE의 기술로만 연구 및 개발된 한국어 **KorPOSTag()**로 제공됩니다. 기존 KOMORAN 1.0 대비 속도, 정확률이 개선되었으며 타 형태소 분석기와 달리 여러 어절을 하나의 품사로 분석 가능함으로써 형태소 분석기의 적용 분야에 따라 공백이 포함된 고유명사(영화 제목, 음식점명, 노래 제목, 전문 용어 등)를 더 정확하게 분석할 수 있습니다.

- ✓ Skills used
 - **javaaddpath** (자바 라이브러리 연동)

Save

Load

Add

POS

☐ OFF

☒ ON

List of Songs

by Popularity

	Year	Title	Artist	Lyrics
1	2014	지친 하루	윤종신	거기까지라고 누군가 툭 한마디 던...
2	2013	내일 할 일	윤종신	이른 아침 일어나야해 내일 우리들...

POS tagging in progress...

21th lyrics tag completed

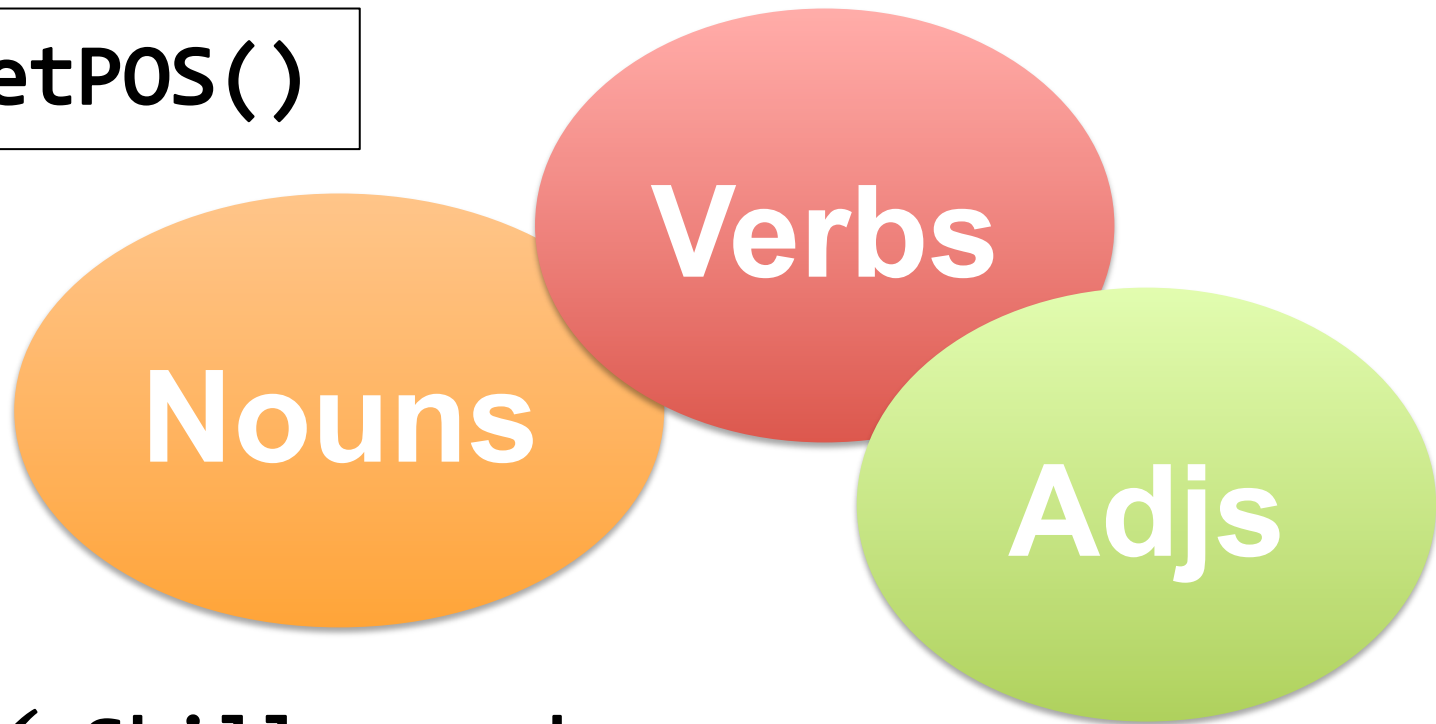
8	2015	The Color	윤종신	어제도 반속 그제도 반속 그럭저럭...
9	2011	늦가을	윤종신	옷장을 열어 보았어 몇 벌이 눈에 ...
10	2014	New You	윤종신	말하지 말고 마구 사랑만 하자 지...
11	2014	Wild Boy	윤종신	난
12	2013	너에게 간다	윤종신	너
13	2013	사랑의 역사	윤종신	수
14	2011	나이	윤종신	안되는 걸 알고 되는 걸 아는 거 그 ...
15	2015	굿나잇	윤종신	굿나잇 하루가 눈꺼풀 위로 내려앉...
16	2010	이별의 온도	윤종신	또 하나의 계절이 가고 찬 바람은 ...
17	2010	본능적으로	윤종신	본능적으로 느껴졌어 넌 나의 사람...
18	2013	굿바이	윤종신	날 생각해봐 처음부터 내가 세상에...
19	1996	환생	윤종신	다시 태어난 것 같아요 내 모든게...
20	2014	Bad Girl	윤종신	벌써 우리도 단팥이데 말하는 거

POS tagging In GUI

2014	쓰담쓰담	10cm	알=VV,고=EC,있=VV,지=...
2015	10월의 날씨	10cm	오늘=NNG,의=JKG,날씨...
2014	스토커	10cm	나=NP,도=JX,알=VV,아=...
2014	그리워라	10cm	가끔=MAG,씩=XSN,자꾸...
2011	사랑은 은하...	10cm	사랑=NNG,은=JX,은하수...
2012	애상	10cm	말리=VV,지=EC,말=VX,...
2015	비밀연애	10cm	잠깐=MAG,만=JX,요=JX...
2011	그게 아니고	10cm	그=NP,게=EC,아니=VCN...
2011	안아줘요	10cm	비오=NNP,는=JX,날=NN...
2010	Good Night	10cm	오늘=NNG,무슨=MM,일...
2014	짝사랑	10cm	그대=NP,의=JKG,표정=N...
2012	Fine Thank...	10cm	너=NP,의=JKG,얘기=NN...
2010	세 번 다시	10cm	가끔=VV,다시=VV,세번=...

(5)Get specific POS

getPOS()



- ✓ Skills used
 - regexp lookahead assertions

'(?<=\w+)=VV' '\w+(?==NP)'

(6) Stopwords removal

Q: Do they mean anything?

1	'내'
2	'그'
3	'난'
4	'날'
5	'수'
6	'내가'
7	'나를'
8	'나'
9	'나의'
10	'너의'
11	'널'
12	'이'
13	'나는'



...???

`stopWords()`

(6) Stopwords removal

- Stopword list?
 - 150 top frequency items in lyricsDB
 - **Too common** words
 - **Removed** via stopWords()
to capture distinct characteristics
of each dataset
- ✓ Skills used
 - **sprintf** (in for loop)

(7) Frequency analysis

FreqAnal()

- Count word or morpheme frequency in each song in data

- ✓ Skills used
 - **unique, countcats** in multiple for loops

	Words	Count
1	싫다	4
2	어지럽다	4
3	그립다	3
4	아름답다	3
5	나쁘다	2
6	두렵다	2
7	뜨겁다	2
8	부끄럽다	2
9	슬프다	2
10	어떨다	2
11	어렵다	2
12	예쁘다	2
13	작다	2
14	차갑다	2
15	괜찮다	1
16	귀찮다	1
17	깊다	1

(8) Wordcloud

wordcloud()

- **Bigger** fontsize for more frequent items
- Pseudo-randomly assign **location** of each words



산울림 (형용사)

끝
추억
그리움
거리
사랑
소리
대답
숨
결
향기
뒤
친구
소리
소리
숨
숨
숨
행복
언제
처음
그
하루
손
미소
한참
아픔
노래
어디
후회
슬픔
고개
잠
이름
전화
얼굴
당신
바보
웃음
시작
아무
바람
걱정
거짓말
입술
침
기대
꽃
마지막
집
그날
곳
간직
자리

윤종신(명사)

타다 나오다 꾸다
달다
데려가다
어쩌다 설레다
떨다 쉼다
교지다 돌다 가지다
만지다
맞추다 놀라다 숨기다
다치다
떨리다 감다 들어가다 꿈꾸다
넘다
다니다 녹다
꺼내다 마주치다
위하다
뛰다 느끼다
그리다
먹다
만들다
넘치다
가다 따라오다
달다 채우다
그려다 울리다
견다
바라보다
놀다
향하다
줄타치다
달리다
서다
들다
버지다
빛나다
따르다
빠지다
밟다
내리다
감추다
담기다
아다

$f(x)$ (동사)

(+) Lyrics Estimator

GUI demo

- web based