

Find Relation Between Singers via Lyrics Analysis

Team: Ave-2013

20120717 최현태

20125753 박준영

20134570 정채진

20155255 김제환

20165953 주광우

Data Crawling: 김재환

Data visualization: 김재환, 정채진

Algorithm Implementation: 주광우, 박준영, 최현태

Evaluation: 정채진

What Is This Project For?

Compare the similarities of individual singers by lyrics.

- Input
 - The lyrics of various singers
 - More lyrics may get better results
- Output
 - Similarity between singers
 - The vocabulary each singer usually uses
 - Representative songs

Idea

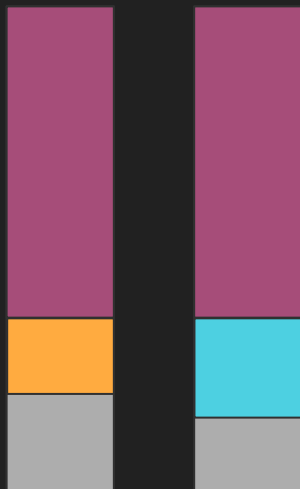
- Singers may have their own favorite vocabulary and expressions
 - Related to genre
 - Reflect their characteristics
- Find **representative terms/songs** and compare through them

Question: How to Choose Data?

- Singers have different # songs
 - More lyrics, more terms
 - More lyrics, higher term frequencies
- Solution
 - Find representative terms
 - Select representative songs

Analysis: Find representative terms

Singer 1 Singer 2



Terms can be separate into

- Daily common words
- Genre common words
- The singer prefers words

Analysis: Find representative terms/songs

- For each singer
 - Find representative terms
 - Use his/her lyrics to build VSM, use tf-idf value as weight
 - Sort each lyrics vector by element weight
 - Chose top K terms of each lyrics as part of the representative terms
 - Select representative songs
 - Count how much representative terms each song contains
 - Use top M songs that contains most representative terms as the representative song

Analysis: Compute Similarity

- Build VSM, compute cosine similarity
 - Different Approaches
 - Keep each song's lyrics in separate files
 - Collect the singer's all lyrics into one single file
 - Add commonly used term into stopword set
 - Use tf-idf weight
 - LSI
 - Approach with best result
 - Collect each singer's all lyrics into as single file, use tf-idf as weight, without LSI

Data Source



We choose the top artists from billboard charts, then crawling the lyrics of each artists.

Select artists from different genres, so we can see the difference between different genre, and difference in the same genre.

Test And Evaluation:

```
dict_keys(['officer-ricky', 'psycho', 'the-realest-niggas', 'who-u-rep-with', 'ghetto-qu-ran-forgive-me-pt-1', 'pt2-bump-heads', 'ghetto-qu-ran', 'how-to-rob', 'banks-victory', 'i-run-new-york', 'that-s-what-s-up', 'how-to-rob-an-industry-nigga', 'who-u-rep-wit', 'i-smell-pussy', 'banks-workout', 'catch-me-in-the-hood', '187-ya-yo', 'rap-game-dvd-live', 'patiently-waiting-feateminem', '8-mile-road-g-unit-remix', 'patiently-waiting'])
```

50-cent funk-flex

결과값 :

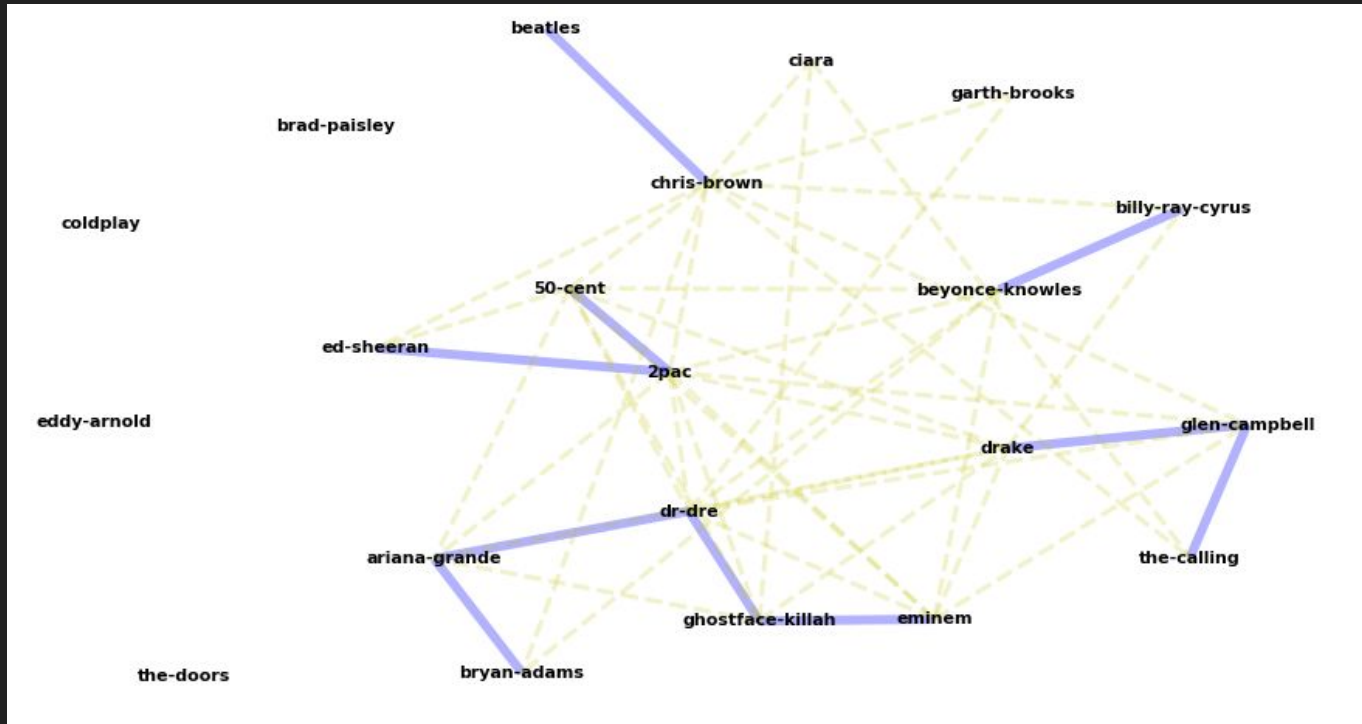
```
artist : 50-cent // similarity : 0.37075907441363154
artist : drake // similarity : 0.28261779440890245
artist : 2pac // similarity : 0.27825511453006874
artist : ghostface-killah // similarity : 0.27807923973101856
artist : dr-dre // similarity : 0.250954910459257
artist : chris-brown // similarity : 0.24501727353660102
artist : eminem // similarity : 0.2302597947892861
artist : ed-sheeran // similarity : 0.1519189400759672
artist : ciara // similarity : 0.15140358885567262
artist : beyonce-knowles // similarity : 0.13857150728997866
artist : bryan-adams // similarity : 0.1285275428744875
artist : ariana-grande // similarity : 0.11902364739316973
artist : brad-paisley // similarity : 0.11695132909924798
artist : the-calling // similarity : 0.11012415455913786
```

Test And Evaluation: Similarity

Most Similar Singers		
Dr.Dre	2 pac	0.184
50 cent	Ghostface-killah	0.163
Eminem	Dr.dre	0.161

Most non-similar Singers		
The Calling	Ciara	0.025
Ariana Grande	Eddy Arnold	0.030
Drake	Coldplay	0.033

Test And Evaluation: Networking



blue > mean + 0.5std

yellow > mean

Test And Evaluation: WordCloud of similar artists



Dr. Dre

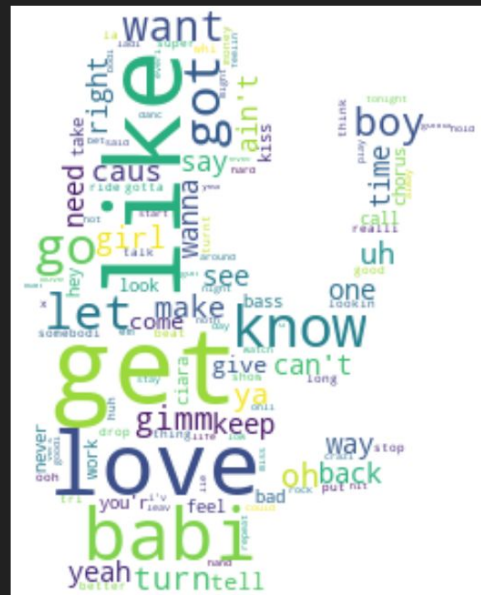


2 Pac

Test And Evaluation: WordCloud of nonsimilar artists

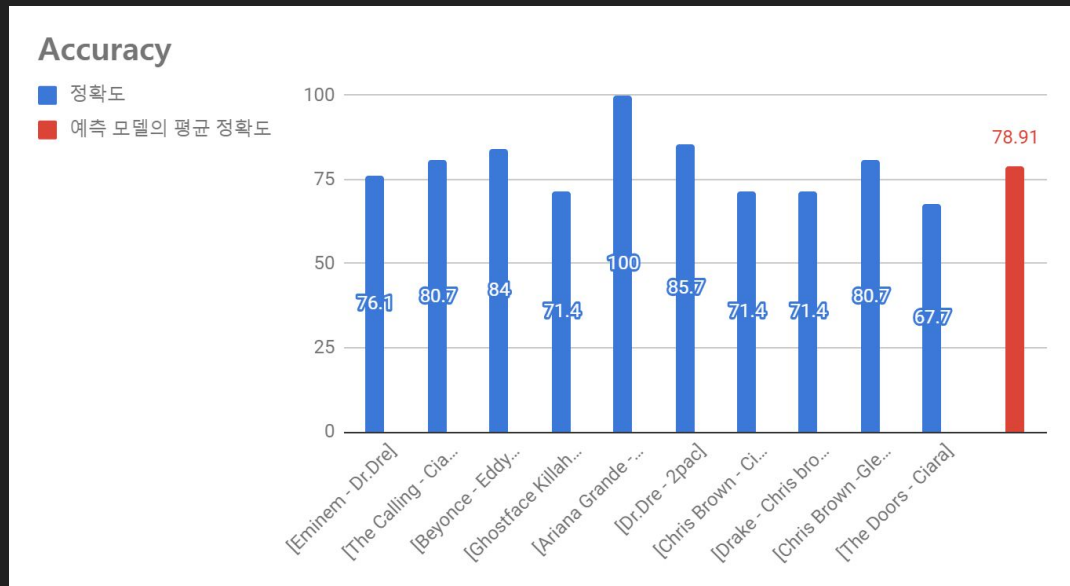


The Calling



Ciara

Test And Evaluation: Survey



- We have 5 similar artists group, 5 non-similar artists group.
- Surveyed by 21 people

Survey link : [link](#)

Spread sheet for result : [link](#)

Thanks