

# **A Cross Domain Lyrics Recommendation from Tourist Spots Reviews with Distributed Representation of Words**

**Yihong Han, Yoko Nishihara (Ritsumeikan University)**

**Ryosuke Yamanishi (Kansai University)**

**Kenta Oku (Ryukoku University)**

Yihong Han : [is0387ps@ed.ritsumei.ac.jp](mailto:is0387ps@ed.ritsumei.ac.jp) Ryosuke Yamanishi : [ryama@kansai-u.ac.jp](mailto:ryama@kansai-u.ac.jp)

# Background



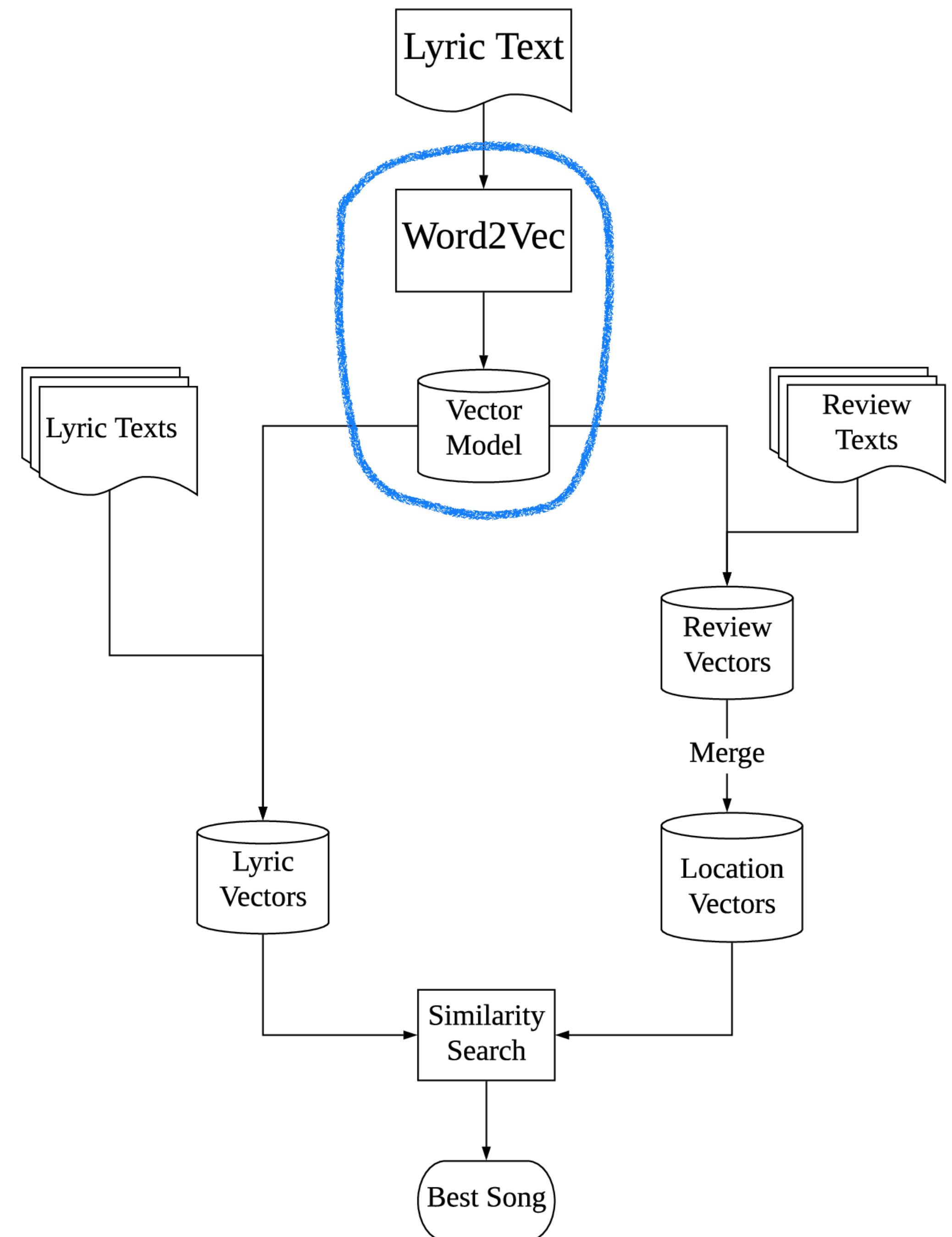
- The listening style has changed a lot during the past decades
  - Subscription service is providing a "unlimited" music access
  - "Do something while listening music" become a ordinary style for us
- What will happen when MUSIC meets TOURISM?
  - It is fantastic for tourists to enjoy music that is similar to the atmosphere of tourist spot they are visiting
  - A method for evaluating the similarity between music and tourist spot is needed

# Method Concept

- The GOAL: To recommend the music which is similar to the atmosphere of tourist spot.
  - Personality of Spots: Reviews on tourism website
  - Personality of Songs: Lyrics of the song
- Swift from reviews to (pseudo) lyric using distributed representation of words
  - The directly comparison between is difficult
  - Distributed representation model is built from lyric corpus
  - Review vectors generated from review texts but using distributed representation model built from lyrics corpus: An idea to deal reviews as **PSEUDO LYRICS**
  - The similarity calculation between lyrics and pseudo lyrics is theoretically possible
- The lyric with the highest similarity to the spot reviews become the recommendation result

# Proposed Method ①

- NEXT → Word Distributed Representations Model
- Quantifying Lyrics and Tourist Spot Reviews to Vectors
- Merging Review Vectors to Spot Vectors
- Lyrics Recommendation Based on Similarity Between Lyrics and Spots

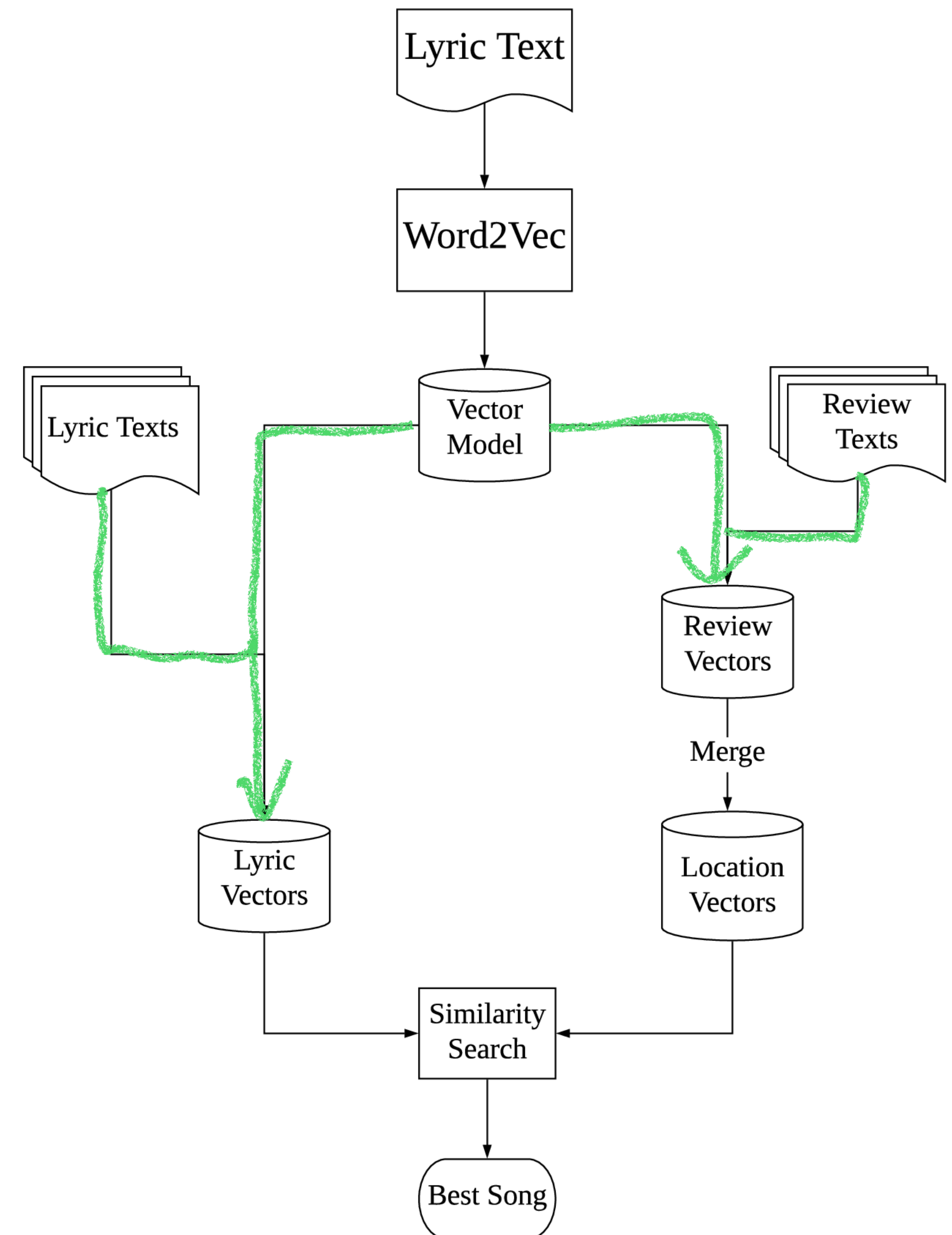


# Word Distributed Representations Model

- Words Distributed Representation: convert words to vectors with vector calculation supported
  - Article vector can be built from word vectors by vector calculation.
  - It is possible to build lyric vectors & review vectors using word distributed representation.
- Word distributed representations model built by Word2vec framework
  - Lyrics corpus of 94451 English songs fetched from [AZlyrics.com](http://AZlyrics.com)
  - Word2vec training parameters: size=300, window=10, min count=2, iter=10

# Proposed Method ②

- Word Distributed Representations Model
- NEXT → Quantifying Lyrics and Tourist Spot Reviews to Vectors
- Merging Review Vectors to Spot Vectors
- Lyrics Recommendation Based on Similarity Between Lyrics and Spots



# Quantifying Lyrics and Tourists Spots Reviews to Vectors

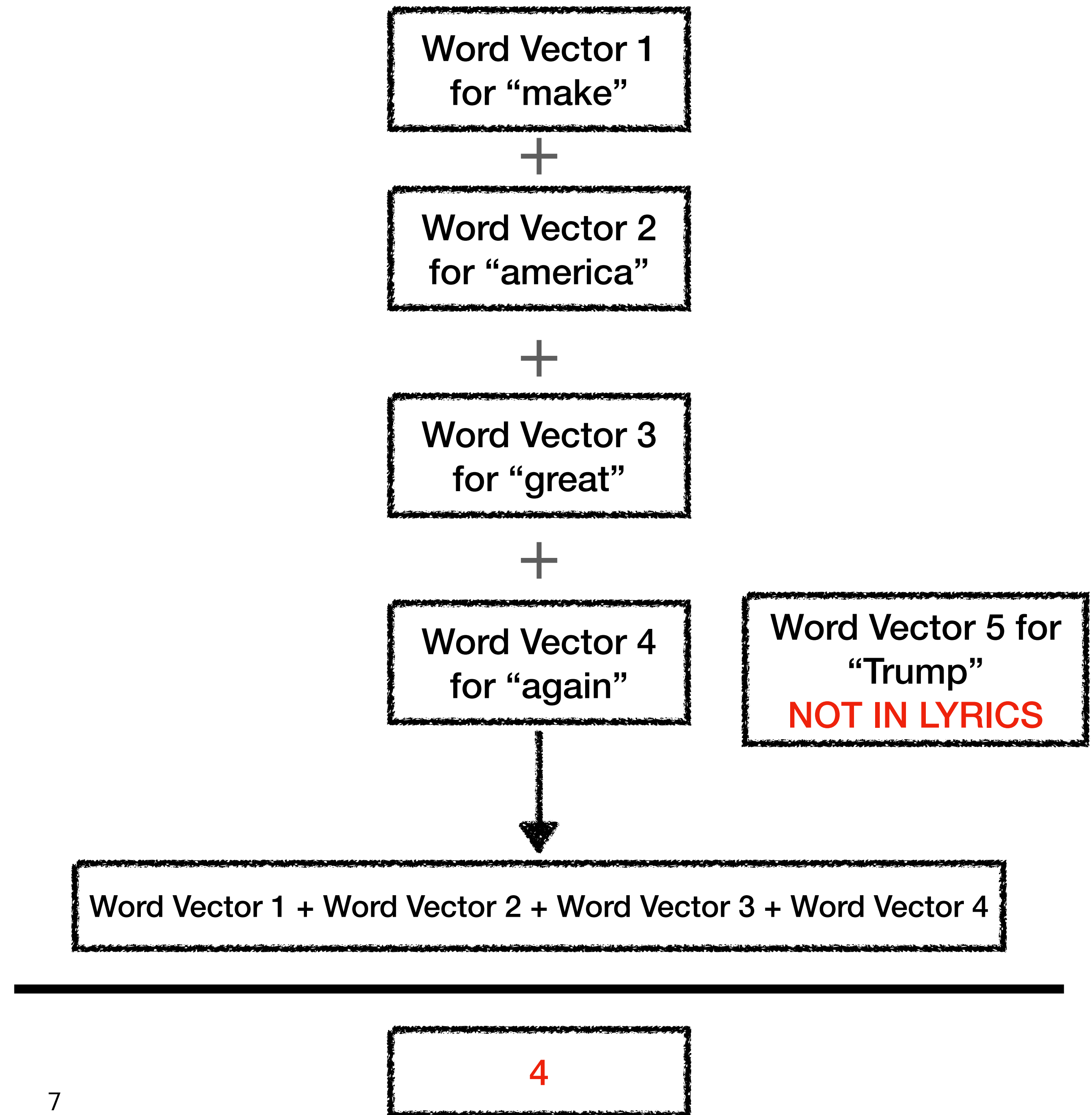
- Fetch word vector from word distributed representations model
- Lyric/Review vector is the mean of all word vectors in its text.

$$\bar{V} = \frac{\vec{v}_1 + \dots + \vec{v}_n}{n}$$

$\bar{V}$  is the vector of a lyric or a spot review

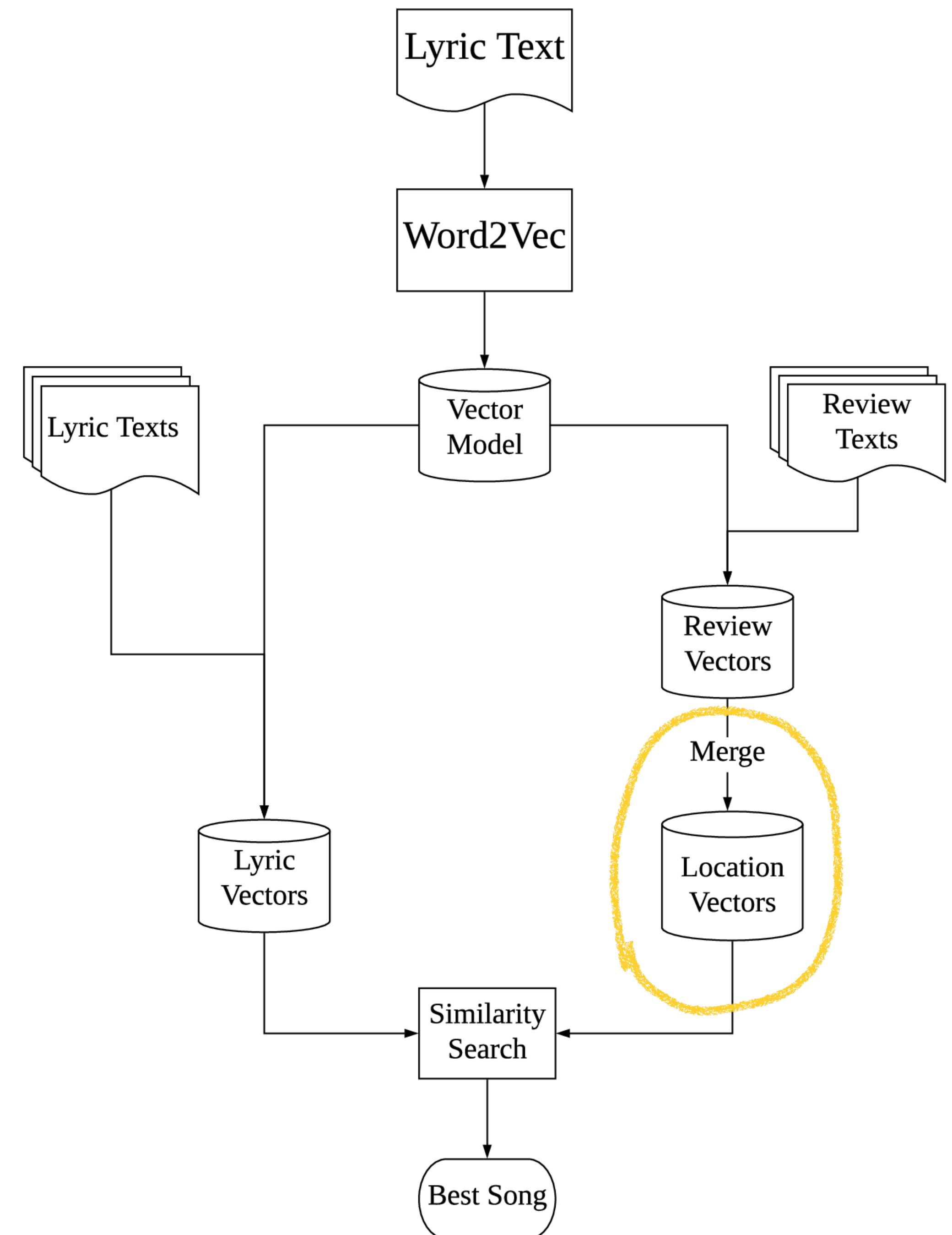
$\vec{v}_n$  is the vector of each word

$n$  is the number of words in a lyric or a spot review



# Proposed Method ③

- Word Distributed Representations Model
- Quantifying Lyrics and Tourist Spot Reviews to Vectors
- **NEXT** → Merging Review Vectors to Spot Vectors
- Lyrics Recommendation Based on Similarity Between Lyrics and Spots





# Merging Review Vectors to Spot Vectors

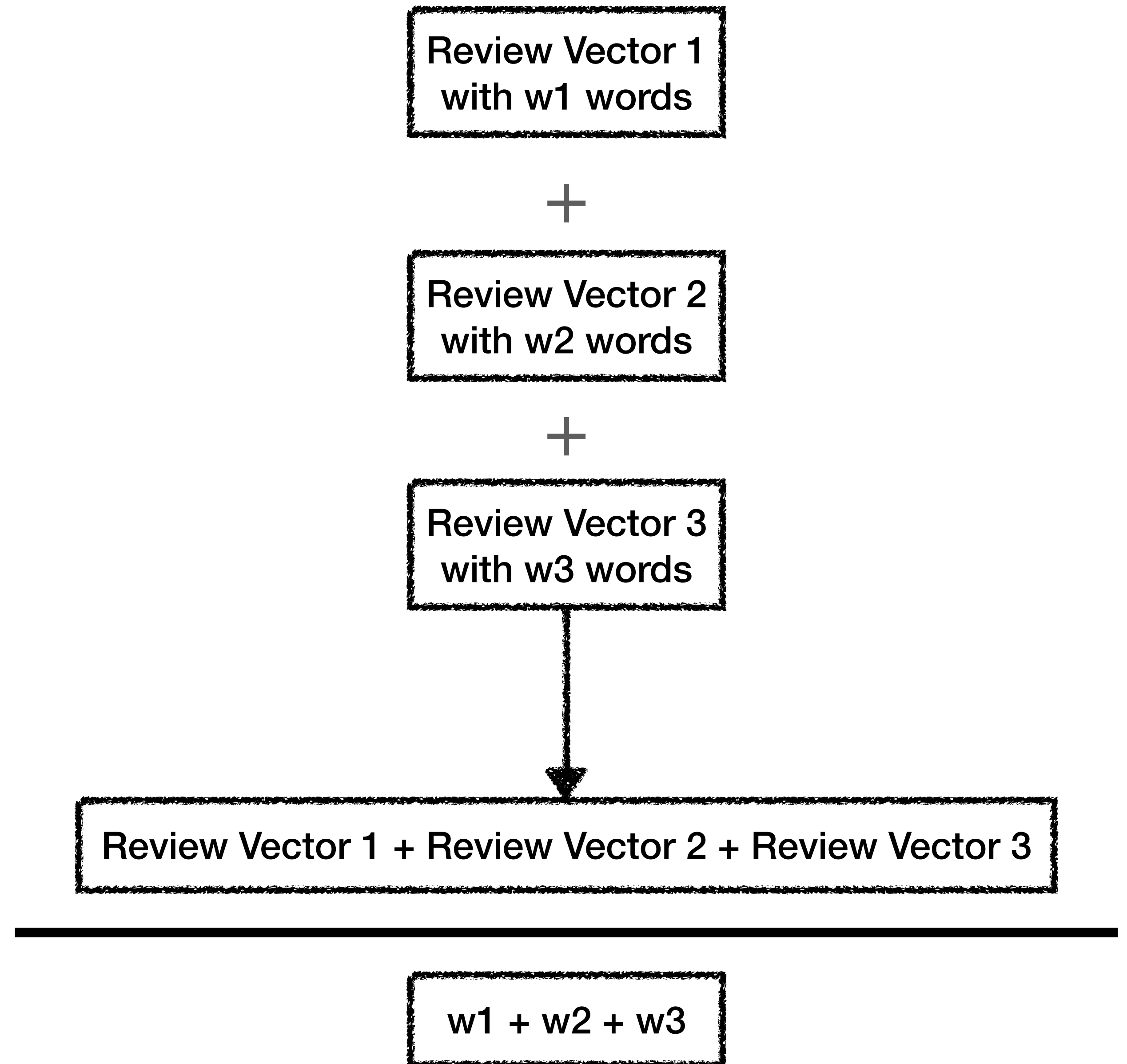
- The individuality and general property of each spot can be extracted by gathering the reviews from each visitor
- Review vectors is merged with a weighted arithmetic mean
- The more words, the more information, the more contribution, the more weight

$$\bar{X} = \frac{\omega_1 \vec{x}_1 + \dots + \omega_n \vec{x}_n}{\omega_1 + \dots + \omega_n}$$

$\bar{X}$  is the vector of each spot

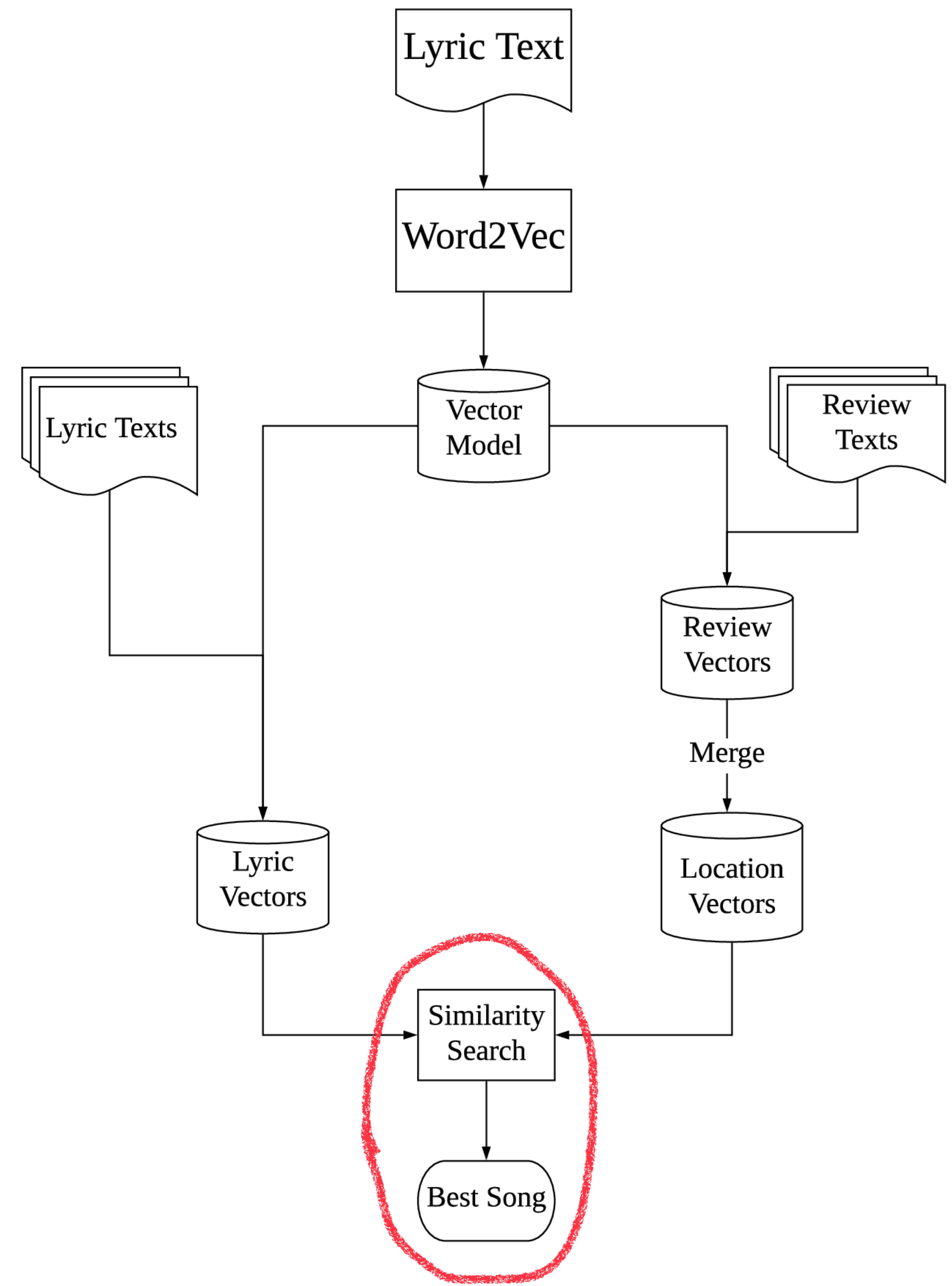
$\omega_n$  is the word number of  $n$ th review

$\vec{x}_n$  is the vector of  $n$ th review



# Proposed Method ④

- Word Distributed Representations Model
- Quantifying Lyrics and Tourist Spot Reviews to Vectors
- Merging Review Vectors to Spot Vectors
- **NEXT**→Lyrics Recommendation Based on Similarity Between Lyrics and Spots



# Lyrics Retrieval Based on Similarity Between Lyrics and Spots

- For one spot, calculate the cosine similarity between this spot vector and every lyric vectors, then find the lyric vector with highest similarity
- The cosine similarity between spot vector and lyric vector is considered to be the emotional distance of the spot and the lyric
- The lyric with the highest cosine similarity to the spot will be the system output as the recommendation lyric for the spot

$$\bullet \cos(\vec{X}, \vec{V}) = \frac{\vec{X} \cdot \vec{V}}{|\vec{X}| \times |\vec{V}|}$$

$\vec{X}$  is the vector of spot

$\vec{V}$  is the vector of lyric

# System Output Sample

- Each lyric has a lyrics ID
- Each tourist spot is matched to the lyric with the highest similarity to the spot
- Several tourist spots correspond to the same lyrics (over corresponding issue)

Tourist Spots	Recommended Lyrics ID
The Montcalm at the Brewery London City	84127
The Beekman A Thompson Hotel	84057
Conservatory Garden	55628
Riverside Park	74814
Hudson River Park	74814
Roosevelt Island	39663
Fort Troon Park	74814
New York Harbor	56837
Franklin D Roosevelt Four Freedoms Park	53520
Long Beach	54217

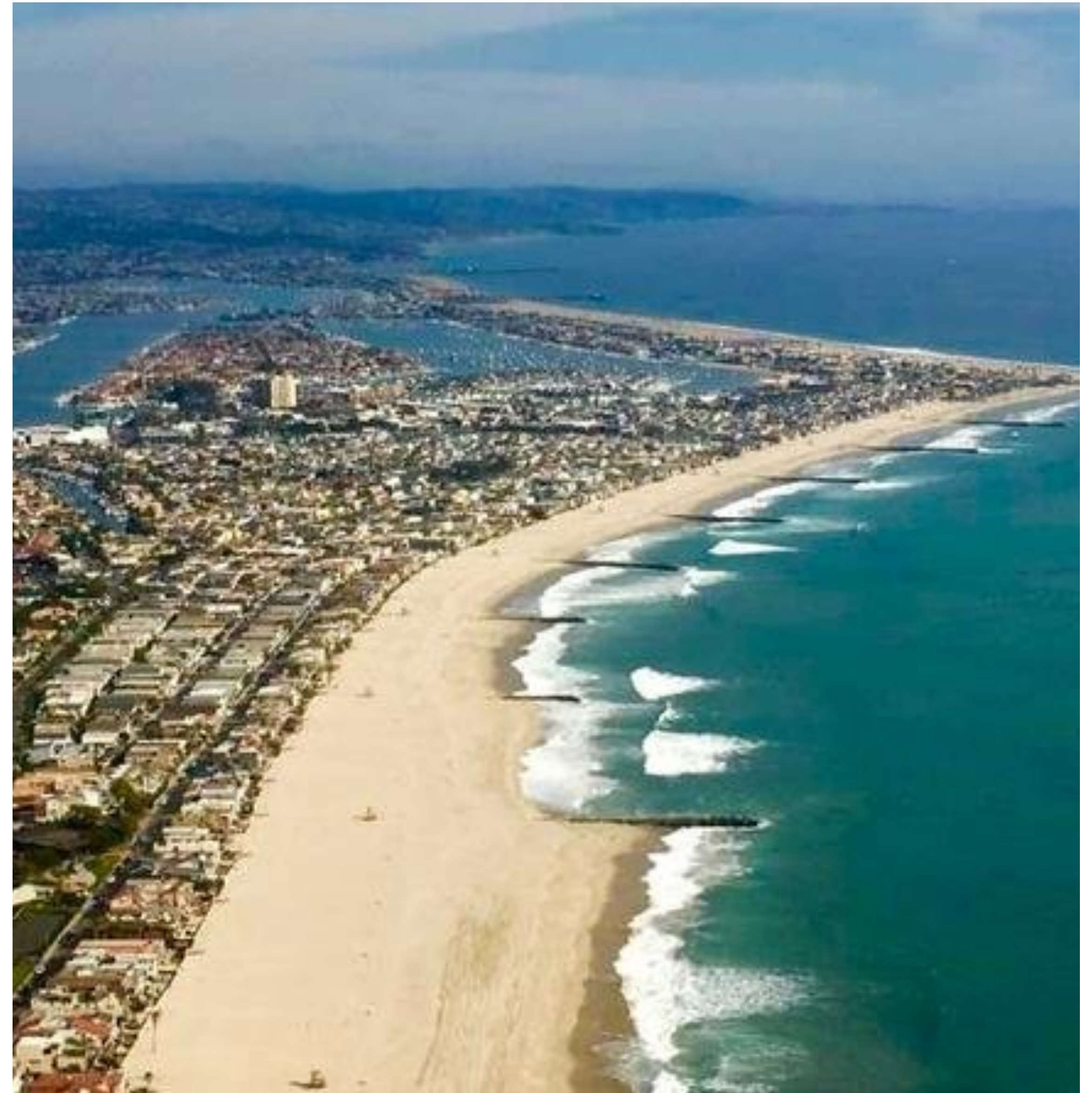
# System Output Example

Long Beach → Miles Between Us (Airplane Song)

A part of the lyric:

When I saw my reflection in her eyes  
That's when I knew that it was time  
I leaned over and kissed her cheek  
And softly said goodbye  
Against the glass I heaved a sigh  
Followed the fading of it's lights  
Watched the plane begin it's flight  
With my eyes into the night  
And I know that I'm a coward  
Who couldn't even tell her  
As she begged me with those eyes  
Give her a reason not to fly  
And it haunts me on the highway  
Headin back to my place  
Why oh why oh why oh why oh why

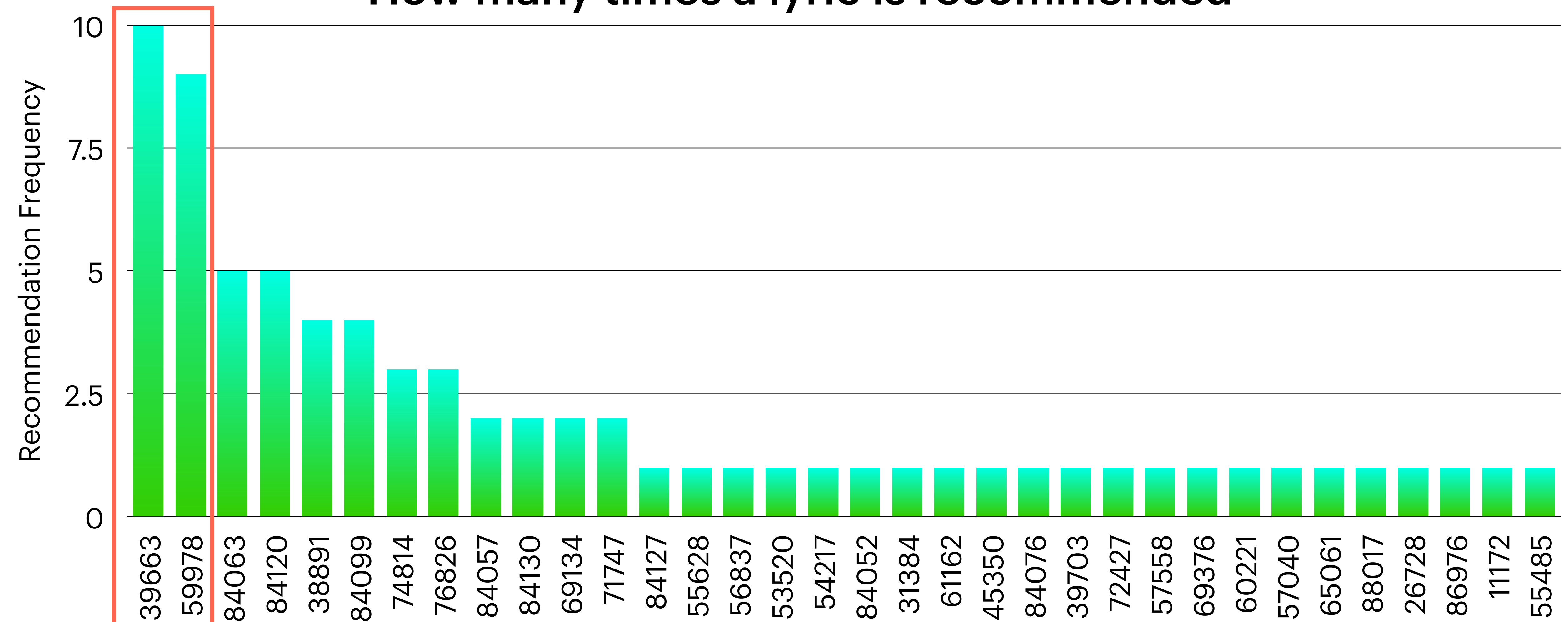
.....



Long Beach in California, USA. Image from [tripadvisor.com](https://www.tripadvisor.com)

# Outline of All System Outputs

How many times a lyric is recommended



over corresponding issue

# Discussing about Over Corresponding Issue

See Through to ID 39663 & 59978

TOURIST SPOTS CORRESPONDING TO LYRIC ID 39663

<b>Tourist Spots</b>	<b>Locations</b>
Roosevelt Island	State of New York, America
Bowling Green	Commonwealth of Kentucky, America
Governors Island National Monument	State of New York, America
SoHo	State of New York, America
West Village	State of New York, America
Meatpacking District	State of New York, America
Twin Peaks	State of California, America
Lincoln Park Conservatory	State of Illinois, America
Greenwich	London, England
Greenwich Park	London, England

TOURIST SPOTS CORRESPONDING TO LYRIC ID 59978

<b>Tourist Spots</b>	<b>Genre of Location</b>
Neue Galerie	Museum
Solomon R Guggenheim Museum	Museum
New York Historical Society Museum Library	Museum
Museum of Arts and Design	Museum
United Nations Headquarters	Organization
Broadway	Street
Radio City Music Hall	Theater
Le Puy du Fou	Theme Park
Westminster	Street

# Discussing about Over Corresponding Issue

## Common Features Between Over Corresponding Locations

- For tourist spots matched with song ID 39663:
  - The common feature maybe the **LOCATION**
  - There are many spots located in the USA, especially in the State of New York
  - “Greenwich” and “Greenwich Park” should be dealt as the same spot
- For tourist spots matched with song ID 59978:
  - The common feature maybe **HISTORICAL**
  - There are many museum in the spots list
  - Radio City Music Hall is a old theater hall with long history
  - Le Puy du Fou is a theme park focus on France history
  - Westminster’s history starts from 16th century