LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics
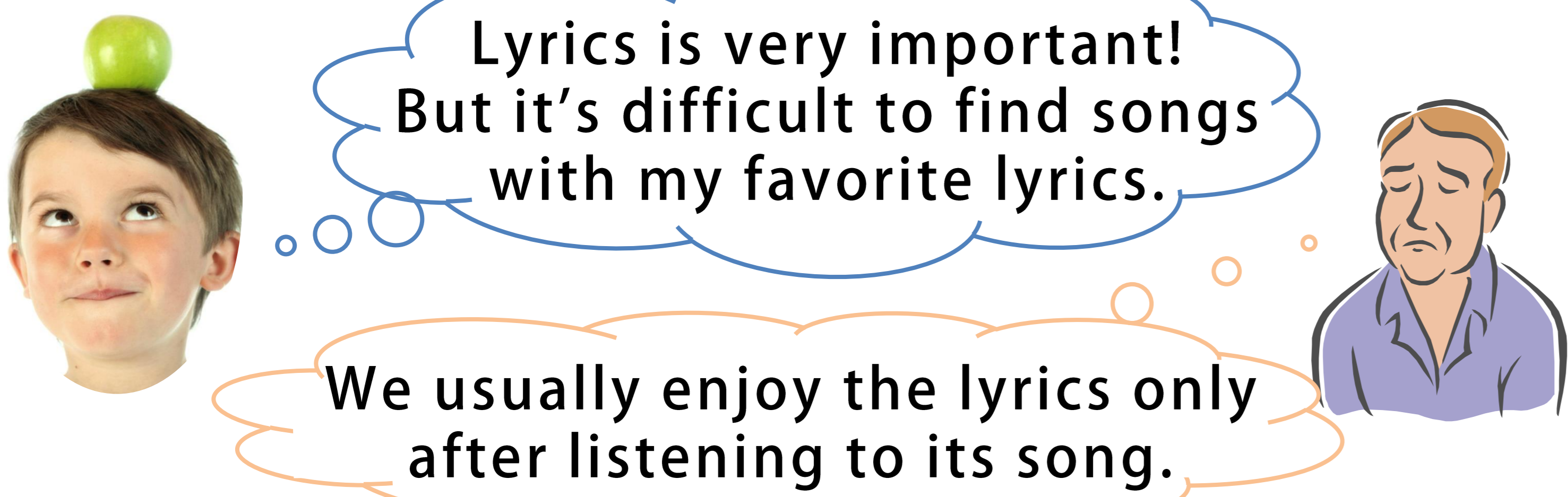
Shoto Sasaki†   Kazuyoshi Yoshii‡   Tomoyasu Nakano*   Masataka Goto*   Shigeo Morishima†

†Waseda University   ‡Kyoto University   *National Institute of Advanced Industrial Science and Technology (AIST)
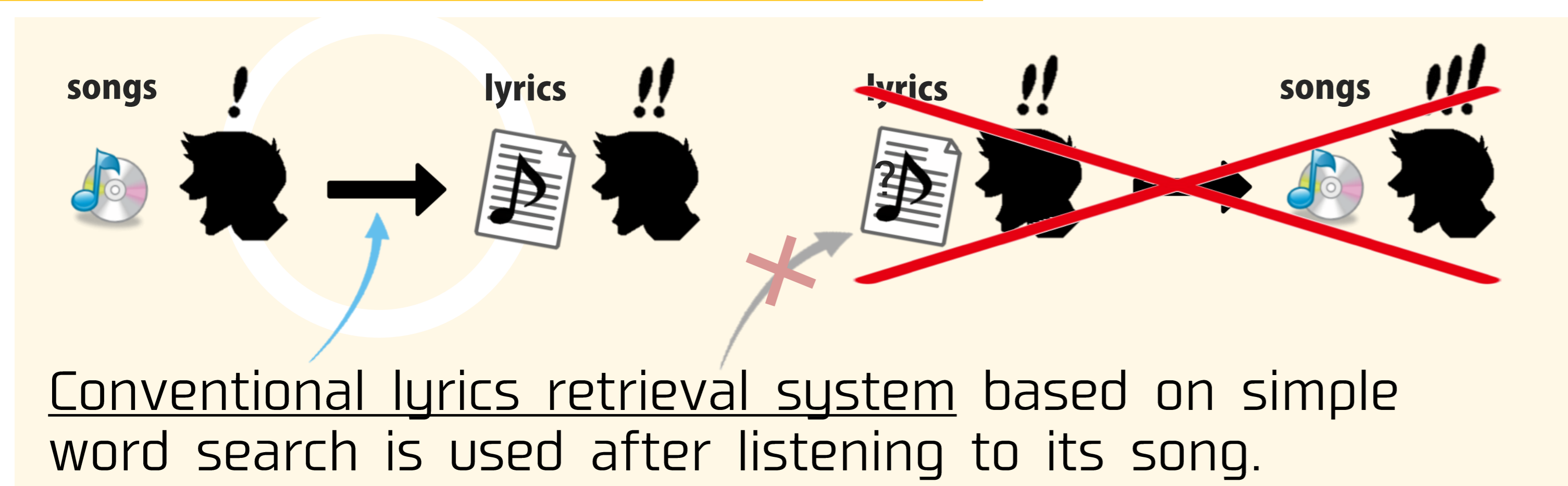
BACKGROUND

Lyrics is very important! But it's difficult to find songs with my favorite lyrics.

We usually enjoy the lyrics only after listening to its song.

GOAL

To assist listeners to encounter songs with unfamiliar but interesting lyrics

lyrics !!

CONVENTIONAL SYSTEM

songs ! → lyrics !! lyrics !! songs !!

Conventional lyrics retrieval system based on simple word search is used after listening to its song.

In addition, those system often fail to reflect user's intention behind a query.

E.g., 'tear'

a drop of salty liquid that comes out of your eye when you are crying
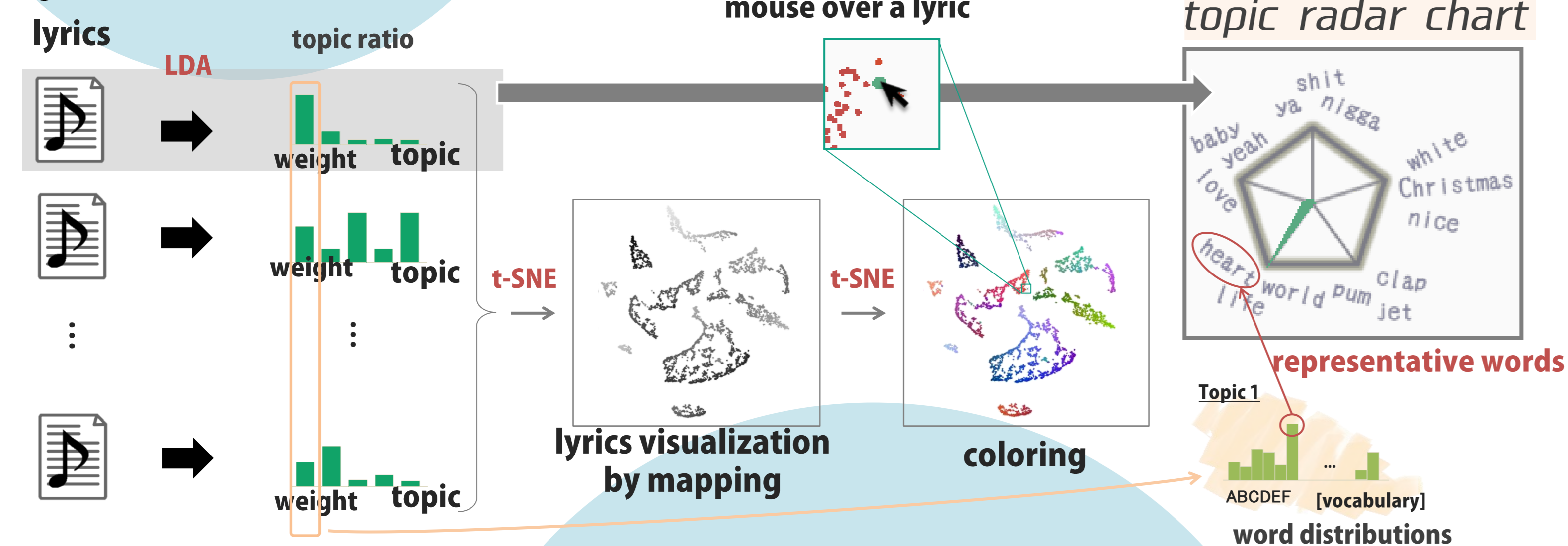
feel heartrending
weep for joy

to damage something such as paper or cloth by pulling it hard or letting it touch something sharp

Longman Dictionary of Contemporary English [http://www.ldoceonline.com/search/?q=tear]

SongMeanings — tear — Results 1 - 20 of 12,724 [http://songmeanings.com/]

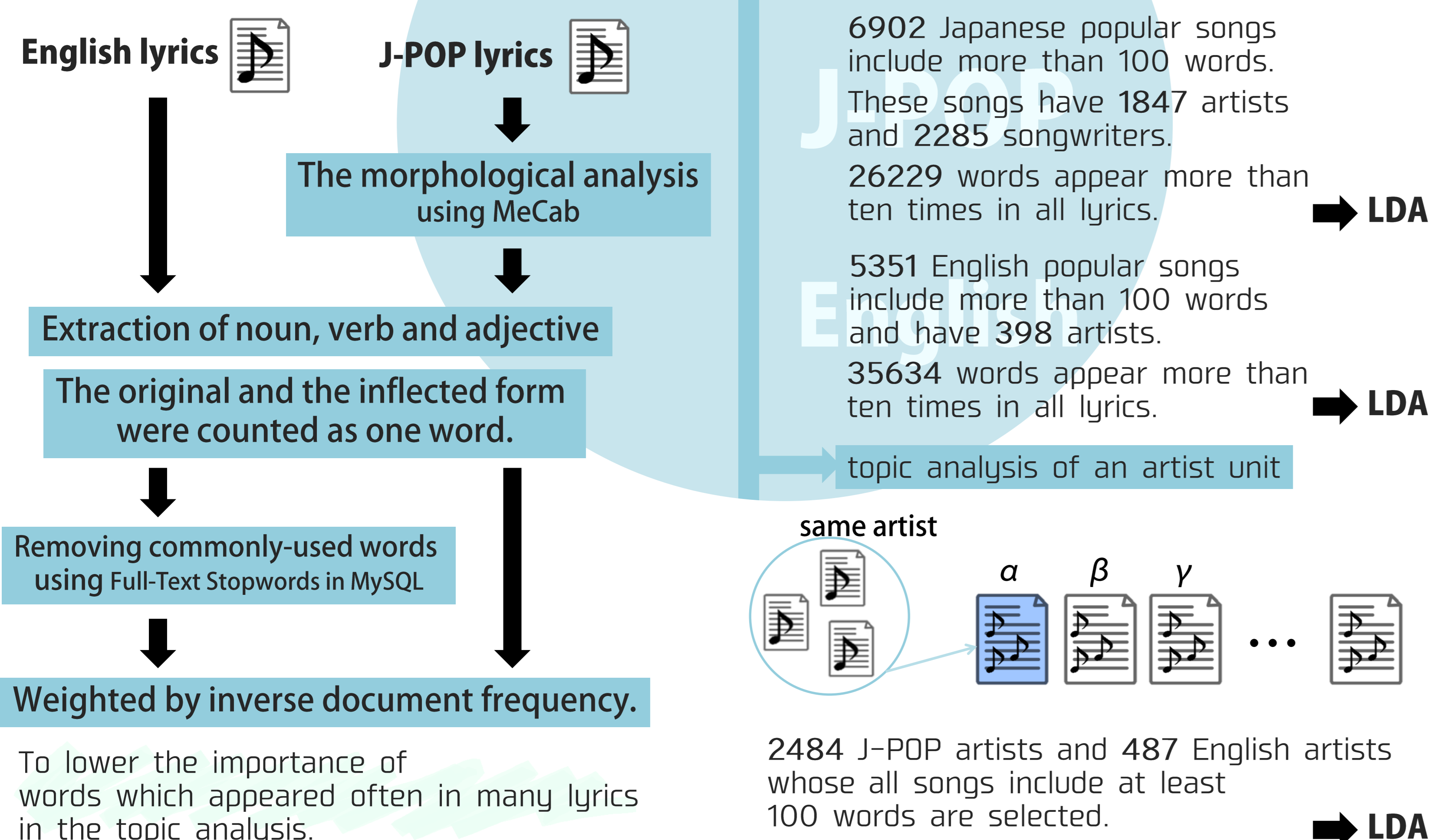| Track	Artist
Second Chance — Please don't cry until tear for me	Shinedown
Otherside — 1 tear it down 1 tear 3 down	Red Hot Chili Peppers
Every Dawg — and there'll probably be a happy tear in my eye but rain or shine	Corey Smith

APPROACH

LyricsRadar

latent topics behind lyrics

analyzes the *lyrics topics* by using latent Dirichlet allocation (LDA) and visualizes those topics to help users find their favorite lyrics interactively.

range of our capacity for processing information
K = 5
perplexity — the number of topics: K
artist unit / lyric unit

The LDA can estimate various *lyrics topics*, five typical topics common to all lyrics in a given database were chosen.

OVERVIEW

lyrics → LDA → topic ratio → weight topic → mouse over a lyric → *topic radar chart*

t-SNE → lyrics visualization by mapping → t-SNE → coloring → representative words

Topic 1 — ABCDEF [vocabulary] word distributions

Training of LDA

English lyrics   J-POP lyrics

The morphological analysis using MeCab

Extraction of noun, verb and adjective

The original and the inflected form were counted as one word.

Removing commonly-used words using Full-Text Stopwords in MySQL

Weighted by inverse document frequency.

To lower the importance of words which appeared often in many lyrics in the topic analysis.

Selection songs include more than 100 words

J 6902 Japanese popular songs include more than 100 words. These songs have 1847 artists and 2285 songwriters.
26229 words appear more than ten times in all lyrics. → LDA

E 5351 English popular songs include more than 100 words and have 398 artists.
35634 words appear more than ten times in all lyrics. → LDA

topic analysis of an artist unit

same artist   α β γ ···

2484 J-POP artists and 487 English artists whose all songs include at least 100 words are selected. → LDA
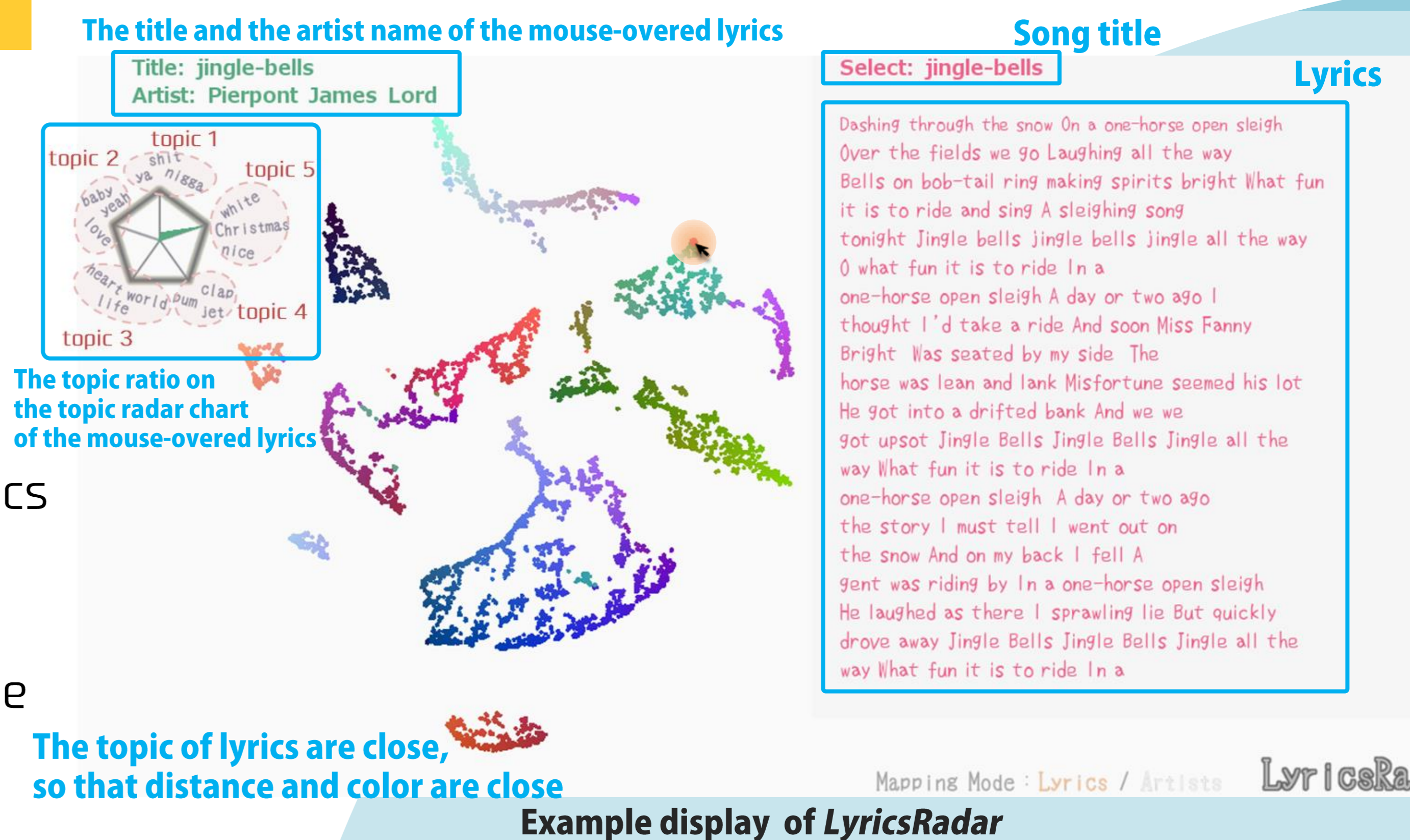
FUNCTIONALITY

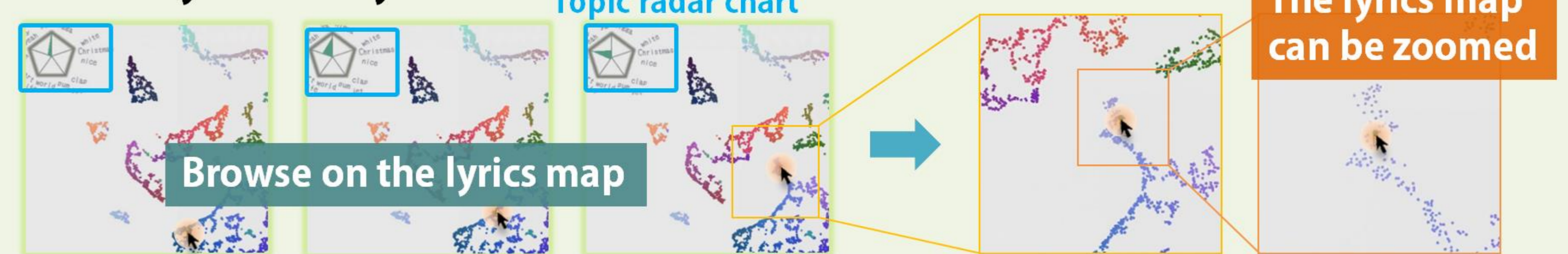LyricsRadar has two visualization functions.

- **the topic radar chart**
  a pentagon shape chart which expresses the ratio of five topics of lyrics

- **the lyrics map**
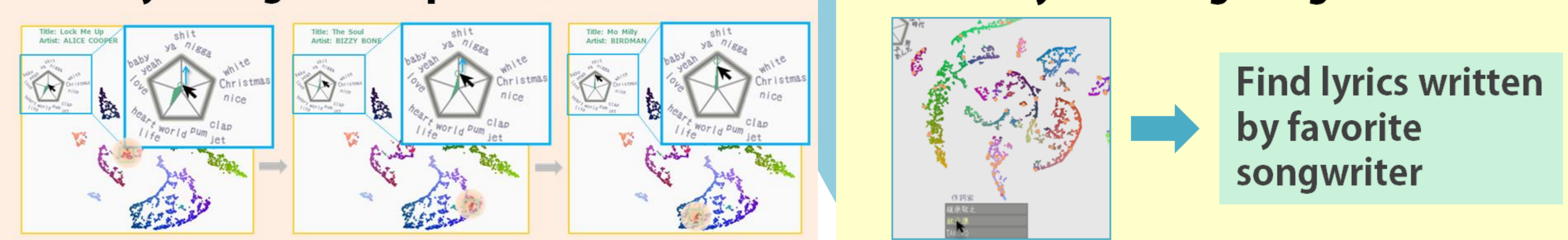 a mapping lyrics dot to the two-dimensional plane

The title and the artist name of the mouse-overed lyrics
Title: jingle-bells
Artist: Pierpont James Lord

Select: jingle-bells   Song title   Lyrics

The topic ratio on the topic radar chart of the mouse-overed lyrics

The topic of lyrics are close, so that distance and color are close

Mapping Mode Lyrics / Artists   LyricsRadar

Example display of *LyricsRadar*

Use favorite lyrics as a query

Song title?   jingle-bells

Find lyrics similar to the query

Browse on the artist map

The size of circle is proportional to the number of songs of an artist

The topic of lyrics are close, so that distance and color are close

Interactively browse lyrics

Topic radar chart

The lyrics map can be zoomed

Browse on the lyrics map

Directly change the topic radar chart

Retrieve lyrics using songwriter

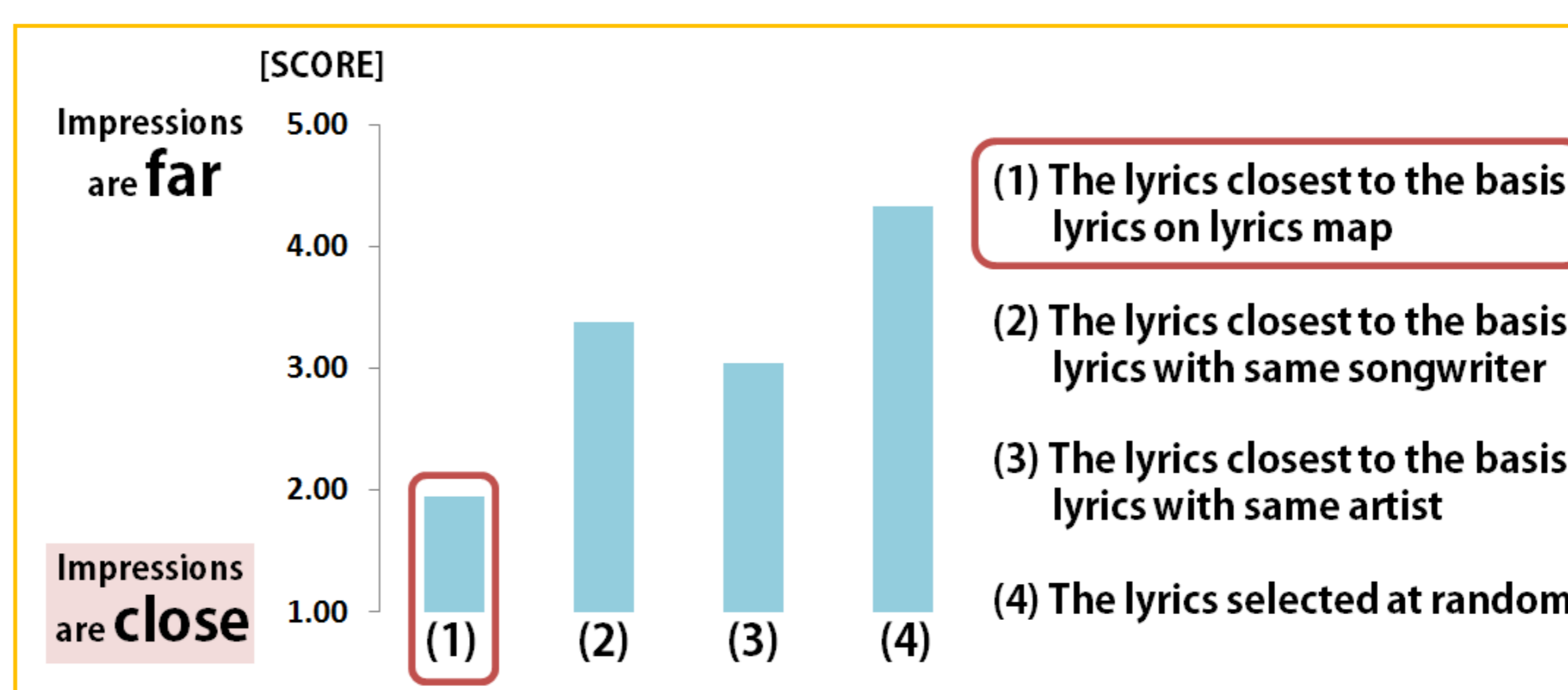**Find lyrics written by favorite songwriter**

EVALUATION

When the lyrics of a song are selected, relative location to other lyrics of the same artist or songwriter in the space is investigated.

The lyrics of a song are selected at random in the space as *basis lyrics* and also *target lyrics* of four songs are selected to be compared according to the following conditions.

We used the results of LDA for the lyrics of the 6902 J-POP songs.

[SCORE]

Impressions are **far** — 5.00, 4.00, 3.00, 2.00
Impressions are **close** — 1.00

(1) The lyrics closest to the basis lyrics on lyrics map
(2) The lyrics closest to the basis lyrics with same songwriter
(3) The lyrics closest to the basis lyrics with same artist
(4) The lyrics selected at random

17 Subjects
all Japanese speakers
age : 21 - 32

Presentation order was random.

The score of (1) was the closest to 1.0, showing our approach to be effective.

As the subjects' comments about the reason of decision, we obtained such responses as a sense of the season, positive-negative, love, relationship, color, light-dark, subjective-objective, and tension.

CONCLUSION

We propose *LyricsRadar*, an interface to assist a user to come across favorite lyrics interactively.

Our main contribution is to visualize lyrics in the latent meaning level based on a topic model by LDA.

Future Work

- the visualization method that can reflect more numbers of topics by keeping an easy-to-use interactivity
- improvements to topic analysis by using hierarchical topic analysis*

* R. Adams et al., Tree-Structured Stick Breaking Processes for Hierarchical Data, Proc. NIPS, 23, pp. 19-27, (2010).