

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 4 No 2 - Desember 2018

Lyric Text Mining Of Dangdut: Visualizing The Selected Words And Word Pairs Of The Legendary Rhoma Irama's Dangdut Song In The 1970s EraTresna Maulana Fahrudin¹, Ali Ridho Barakbah²¹ Universitas Narotama Surabaya, Surabaya² Politeknik Elektronika Negeri Surabaya, Surabayatresnamf@gmail.com¹, ridho@pens.ac.id²**Kata Kunci**

Pemilihan Kata
Pasangan Kata
Lirik
Text Mining
Visualisasi

Abstrak

Dangdut merupakan aliran musik baru yang dikenalkan oleh Rhoma Irama, musisi populer Indonesia yang merupakan penyanyi dangdut legendaris pada era 1970-an hingga sekarang. Ekspresi lirik lagu Rhoma Irama memiliki tema-tema seputar tentang manusia, jalan hidup, cinta, hukum dan hak asasi manusia, tradisi, persamaan sosial, dan pesan-pesan Islam. Namun menariknya, lirik lagu yang dituliskan oleh Rhoma Irama pada tahun 1970an lebih banyak pada tema lagu percintaan. Untuk membuktikan hal tersebut, maka perlu adanya identifikasi pada lagu-lagu tersebut melalui beberapa pendekatan berupa eksplorasi pemilihan kata dan keterhubungan pasangan kata. Jika lirik lagu Rhoma Irama diidentifikasi dalam bidang text mining, ekstraksi teks lirik lagu tersebut akan menjadi pola pengetahuan yang menarik. Kami mengumpulkan lirik lagu dangdut dari sumber web sebagai dataset, lalu kemudian kami telah melakukan ekstraksi data untuk menyimpan komponen lirik termasuk setiap bait dan baris lagu. Kami berhasil menerapkan visualisasi dari frekuensi kata yang paling sering muncul dalam bentuk bar chart, word cloud, term frequency-inverse document frequency, dan network graph. Hasilnya, terdapat pasangan kata yang sering digunakan oleh Rhoma Irama dalam menulis lagunya antara lain cinta-hati (19 baris), hati-rindu (13 baris), hati-sayang (12 baris), cinta-sayang (12 baris), cinta-rindu (11 baris).

Keywords

Selected Words
Word Pairs
Lyrics
Text Mining
Visualization

Abstract

Dangdut is a new genre of music introduced by Rhoma Irama, Indonesian popular musician who was the Legendary dangdut singer in the 1970s era until now. The expression of Rhoma Irama's lyric has themes of the human being, the way of life, love, law and human right, tradition, social equality, and Islamic messages. But interestingly, the song lyrics were written by Rhoma Irama in the 1970s were mostly on the love song themes. In order to prove this, it is necessary to identify the songs through several approaches to explore the selected word and the relationship between word pairs. If each Rhoma Irama's lyric is identified in text mining field, the lyric text extraction will be an interesting knowledge pattern. We collected the lyric from web were used as datasets, and then we have done the data extraction to store the component of lyric including the part and line of the song. We successfully applied the most word frequencies in the form of data visualization including bar chart, word cloud, term frequency-inverse document frequency, and network graph. As a results, several word pairs that often was used by Rhoma Irama in writing his song including heart-love (19 lines), heart-longing (13 lines), heart-beloved (12 lines), love-beloved (12 lines), love-longing (11 lines).

1. Introduction

Dangdut song studies in the fields of linguistics, art and music, social and law have been extensively reviewed in several theoretical

and empirical studies. However, with the development of information technology, song studies in the field of music can be combined with the field of computer science, namely text mining. We have difficulty finding similar research that

addresses the topic of text mining on dangdut songs, therefore, this research is our first research in the field of dangdut songs that tried to explore what can be found through text mining. The role of text mining in this research is one way to visualize the selected words and relationship between word pairs from a set of process to collect dangdut song lyrics and data preparation. We analyze that the theme of the dangdut song was written by songwriters like Rhoma Irama changes every decade. In the first year, in the 1970s, Rhoma Irama wrote more songs about love song themes, and at that time dangdut song lovers in Indonesia really enjoyed, and even memorized each title and lyrics to date. In this research, we also explained historically related dangdut songs to introduce to the public who did not know it, and then discussed technically text mining to explore and visualize it.

The term "dangdut" was first introduced by Billy Silabumi in his short story on *Aktuil Magazine* in 1972 [1]. This magazine also popularized the term "dangdut" to replace the term "Malay Orchestra" or in Indonesian well known as "Orkes Melayu". The word "dangdut" is an idiom of the actual word by Billy Silabumi used to "mock" the Malay Orchestra which is from a musical perspective seems monotonous by only exploiting the sound of "dhang" and "dhut".

The word "dangdut" became popular, the Malay Orchestra musicians not only did not accept the word but also did not reject it. Many musicians were still using the Malay Orchestra to name their group but used the word "dangdut" to refer to the type of music at the time [2]. Therefore, the name "dangdut" comes from onomatopoeia (the formation of a word from a sound associated with what is named) of the traditional musical instruments "gendang or kendang" (a set of two traditional drums) that sounds like "dhang" and "dhut".

Dangdut is a music genre that combines the arts of Malay, Indian, Arabic music with elements of popular Western music such as America and Europe. Dangdut has also developed during the early 1970s in Jakarta which is Metropolitan city and Indonesia Capital City. Even dangdut help the people to make common culture which is caused by multi-ethnic, multi-cultural, multi-lingual of Jakarta in common language [3]. For example in 1970s era, the combination between traditional musical instruments "gendang" and modern musical instruments "drums" has made new dangdut genre. The combination of musical instruments has become creative together with electric guitar, bass guitar, electronic keyboard, and "suling" (flute). The increasing number of musical variations also increased the appearance of famous dangdut singers in the 1970s.

Rhoma Irama is one of the famous singers in 1970s who was called "Raja Dangdut" (the King of Dangdut) by Indonesian people. Rhoma Irama

is the most important artist in the rapid advancement of dangdut music because succeed to revolutionize music through live concerts and recordings. The expression of Rhoma Irama's lyric has themes of the human being, the way of life, love, law and human right, tradition, social equality, and Islamic messages which makes enjoy music lovers[4]. The slogan of dangdut is "Musik Rakyat" (the music of people) was popular among Indonesian people in 1970s [3][5]. Rhoma Irama has written beautiful lyrics on the song entitled "Malam Terakhir" (1972), "Darah Muda" (1975), "Cuma Kamu" (1975), "Gitar Tua" (1977), writing a song until now, and also formed "Soneta Group" which consists of guitarist, mandolinists, two keyboardist, flutist, bassist, rhythmist, and tambour (gendang player). Figure 1 illustrated Rhoma Irama is holding a guitar [4].

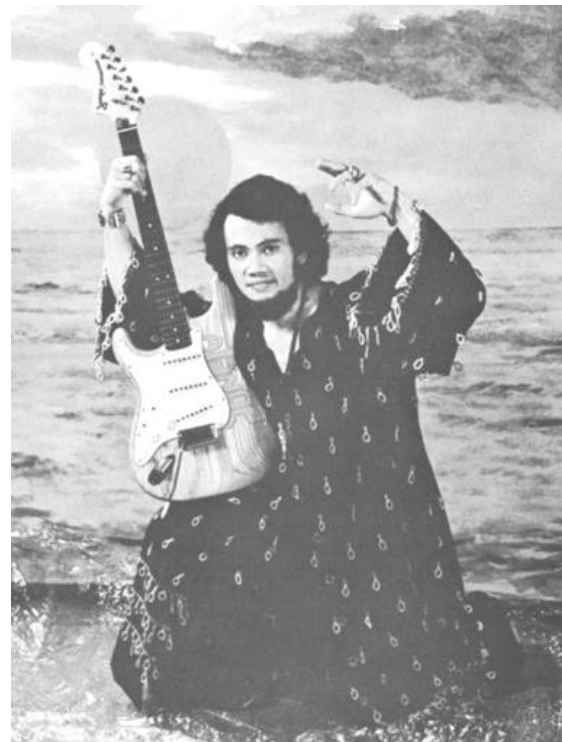


Figure 1. Rhoma Irama, who is the King of Dangdut holding a guitar [4]

Indonesian people know that Rhoma Irama writes the beautiful song lyric for them who listen to dangdut. Each lyric was written by Rhoma Irama has a deep meaning for the listener. If each lyric is identified by computer science especially in text mining or text processing field, the lyric text extraction will be an interesting knowledge pattern. Usually the part of song lyric consists of four lines, however, there is consists of more than four lines and has a letter suffix pattern. The structure of song lyric part as same as paragraphs in text, each part of song lyric separate the topics or ideas that will be expressed in a song. Therefore, this paper will identify the characteristic of Rhoma Irama's dangdut song

lyric to open what the most important words in these song using lyric text mining. The identification begin how to a.) obtain the data collection for preparing data from web sources, b.) split whole of song lyric up into per part of the song lyric and per line of the song lyric part for obtaining all lines of the song, and c.) count the most word frequencies in the form of data visualization (bar chart, word cloud, tf-idf, and network graph).

2. Related Works

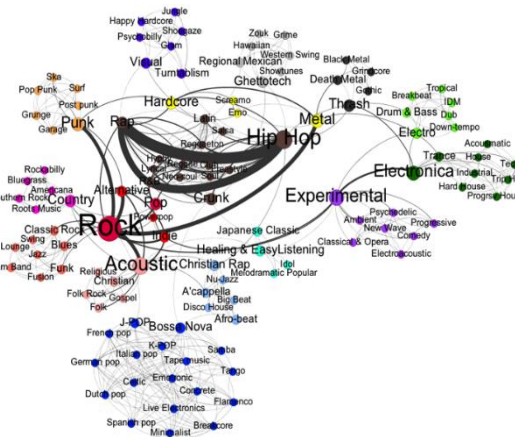


Figure 2. Network visualization of intercorrelation among music genre communities [6]

Many music genres in the world which is each country has its own musical characteristic. The complexity of many musical genres in the world makes a difficulty to visualize their relation. Figure 2 illustrated the network-genre to display a simple visual of grouping each music genre in the cluster [6]. Each music genre is defined by node colour, but there are three strongest music genre communities. Cluster A represents Rock 'n' Roll genres (**warm colours:** red, pink, yellow, orange), Cluster B represents Underground genres (**cold colours:** grey, green, blue, purple), and Cluster C represents African-American-Latino Hip Hop genres (**brown colours**). How many music genres can be conducted research in the world, while dangdut is also a new music genre that was combined between traditional musical instruments using “gendang” sound like “dhang” and “dhut” from Indonesia and modern instrument musical using the electronic keyboard, guitar, and bass from Western. Therefore, to find out more about the trend of text mining research in song or music fields, it needs to know previous research from the other researcher about their topic that was published.

Braja Gopal Patra, et al from Jadavpur University, Kolkata, India. Their research is about mood classification of Hindi songs based on lyrics [7]. Their paper proposed about how to make a recommendation system for playing 461 Hindi lyric songs based on two perspectives. First, they used mood taxonomy to classify the song lyric

based on listener from audio which is inspired from Russel’s circumplex model, there are five mood classes that consist of **Class_Ex** (Excited, Astonished, Aroused), **Class_Ha** (Delighted, Happy, Pleased), **Class_Ca** (Calm, Relaxed, Satisfied), **Class_Sa** (Sad, Gloomy, Depressed), and **Class_An** (Angry, Alarmed, Tensed). Second, they used polarity to classify the song lyric based on the reader which consists of positive and negative, and also to keep consistency of each song lyric that contains multiple emotion or mood labels.

Chutimet Srinilta, et al from King Mongkut’s Institute of Technology Ladkrabang, Thailand. Their research is about lyric-based sentiment polarity classification of Thai songs [8]. Their paper explained how to make the assumption of two classes in sentiment analysis, which are positive lyrics represent a “happy” song while negative lyrics represent a “sad” song. Their data collection is obtained from Chord Café website which provided lyrics and chords of Thai songs consist of 34 groups based on the emotion of **first love, love forever, lonely, painful, broken-heart, and cheerful**. These songs are grouped into two classes include 427 “happy” and 317 “sad” songs. Neural Network Multi-Layer Perceptron (MLP) was applied to classify sentiment polarity of Thai song lyrics. The characteristic of Neural Network is very well to solve Thai song classification problem because the labelling of these song based on input by the human that is subjective and noisy.

3. System Design

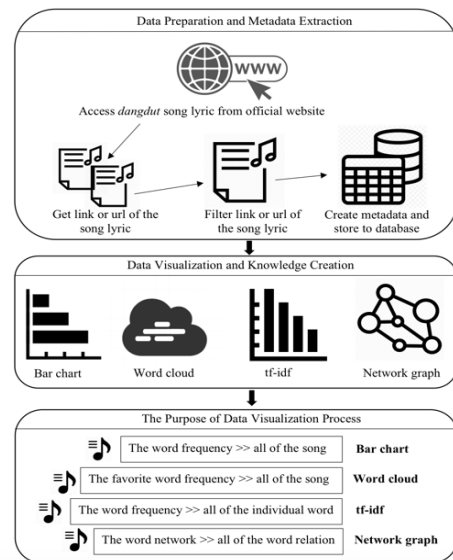


Figure 3. The system design of dangdut song visualization using lyric text mining

In Figure 3 shown the system design of this research including data preparation and metadata extraction process, data visualization and knowledge creation. Data preparation is a step to search the available website which it provides dangdut song lyric and also it ensures the website's source code is easy to read.

Metadata extraction is a step to store all elements of descriptive data to the database. Data visualization in this paper is a way to help understand dangdut song lyric as data into a visual context which consists of a bar chart, word cloud, *tf-idf*, and network graph. Knowledge Creation is an output from process sets to give a representation of any words that are often used by dangdut songwriters, and also the representation of word network in song lyrics.

3.1 Data Preparation

The web page source that provides dangdut song lyric usually has an HTML structure. The syntax `` can easily get the song lyric link based on the available link in per web page. *Lirik Lagu Indonesia* is an official website that provides dangdut song lyrics which actually not only provide dangdut song genre but also it provides the other song genre such as the western song, kid song, and pop song. This website is very easy to be used because there are two pages only to go to the singer page and singer's song lyric page. For example, to access the singer page like "Rhoma Irama" can go to <https://liriklaguindonesia.net/rhoma-irama>, this page will show the song titles list of Rhoma Irama but not the contents of the lyrics. While to access the song lyric page of Rhoma Irama can go to <https://liriklaguindonesia.net/rhoma-irama-pedih.htm>, this page will show the contents of full song lyrics.

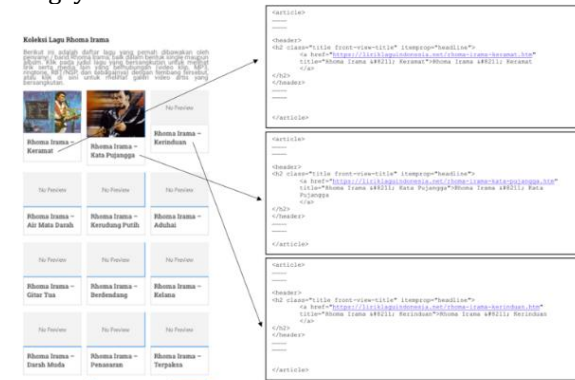


Figure 4. The title list page of Rhoma Irama's dangdut song (without contents of the lyrics) on Lirik Lagu Indonesia official website

The singer page of *Lirik Lagu Indonesia* official website shows there are a total of 12 dangdut song titles listed on per page. Per title box in this page represents the dangdut song title that has information about the singer name and the song title name. Per title box also has an embedded link which will be directed to the lyric contents page. Figure 4 illustrated a view page sources on the list of Rhoma Irama's dangdut song title. The HTML code of this page has the same title box structure with another title box. All link of the box will be collected on the gathering information process.

After the user clicks the title box, *Lirik Lagu Indonesia* official website will redirect this page to another page that is the lyric contents page. This page will show the part of the song lyric that consists of two until four lines on each part. Usually, one part of the song lyrics are repeated in the other song lyrics. Figure 5 illustrated 10 part of the song lyric which is some parts of song lyrics are repeated. If the page source of this page is opened, it will show the separator of each part of the song using one blank line. The one blank line in HTML code using `
`, while new line using `<p>...</p>`. All new line and a blank line will be used for identifying the line position of a song lyrics.

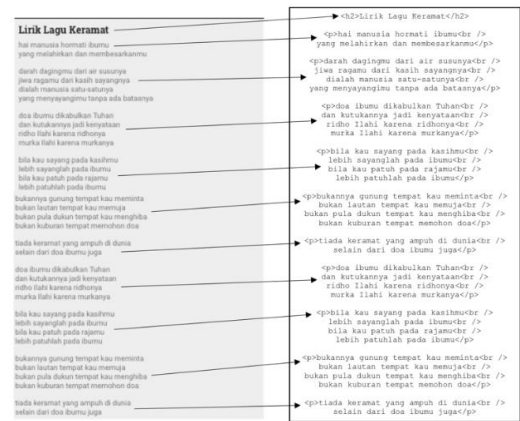


Figure 5. The lyric contents page of Rhoma Irama's dangdut song on Lirik Lagu Indonesia official website

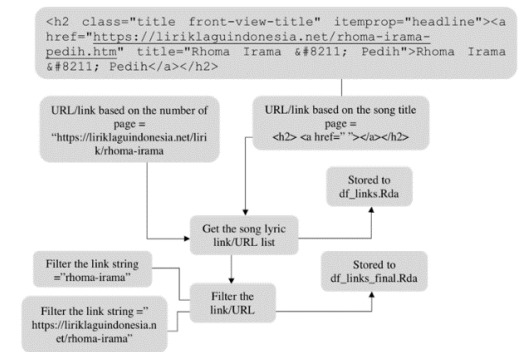


Figure 6. The data preparation to get the link/URL of dangdut song lyric from website

The data preparation is illustrated in Figure 6 about how to collect the dangdut song lyric link from a website and then it is stored in a database. There are two processes consist of **getting the song lyric link/URL list** and **filtering its link/URL** in the following explanation below:

- **Getting the song lyric link/URL list:** A website usually has HTML structure which it can be viewed to page source, the page is written in HTML code `<h2></h2>`. *Lirik Lagu Indonesia* as one of the official website of song lyric provider is also using HTML code. There are two data

inputs which are the first data from the **source** (URL/link based on the number of page:https://.../[singer_name]/page/1;/page/2; /page/3) and the second data from the **link** (sub of singer page as song title: https://.../[singer_name]-[song_title]). Code `<h2></h2>` represents the second level heading in an HTML document to get the title of post page, while code `` represents the anchor tag and hypertext references that are used to identify singer and song title page through link/URL. After the source and link of dangdut song were accessible, it all data will be stored in a database which is called *.Rda (RData format). It is not fully database, but this data format is allowed to save the data structured in data frames, matrices, and vectors. The Figure 7 is illustrated 23 data of source and link which are represented in URL, but they are not filtered. These source and link data still contain URL that is not related to Rhoma Irama's dangdut singer. The red circle in Figure 7 shows some unrelated song links.

link	source
https://liriklaguindonesia.net	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-keramat.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kata-pujangga.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kerinduan.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-ai-mata-darah-irama	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kerudung-putih.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-aduhai.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-gitar-tua.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-berdandang.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kelana.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-darah-muda.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-penasaran.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-terpaksa.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-cuma-kamu.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-any.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-pedih.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-malam-terakhir.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-setetes-air-hina.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-syahdu.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/press-release-rhoma-irama-sore...	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/agrikulture-lagu-cinta-cintaan.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/sule-aku-cinta-indonesia.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2

Figure 7. The source and link of dangdut song are represented in URL (unfiltered)

link	source
https://liriklaguindonesia.net/rhoma-irama-keramat.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kata-pujangga.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kerinduan.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-ai-mata-darah-irama	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kerudung-putih.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-aduhai.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-gitar-tua.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-berdandang.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-kelana.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-darah-muda.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-penasaran.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-terpaksa.htm	https://liriklaguindonesia.net/lirik/rhoma-irama
https://liriklaguindonesia.net/rhoma-irama-cuma-kamu.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-any.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-pedih.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-malam-terakhir.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-setetes-air-hina.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2
https://liriklaguindonesia.net/rhoma-irama-syahdu.htm	https://liriklaguindonesia.net/lirik/rhoma-irama/page/2

Figure 8. The source and link of dangdut song are represented in URL (filtered)

- **Filtering link/URL:** 23 links list must be filtered which they must be related to the item of "rhoma-irama". The filtering process of these link can be done in two ways, the first is **direct string filter** and the second is **URL address filter**. The direct string filter is

identifying the structure of link based on item string, for example, if there is item string filter "rhoma-irama", then all of the links which they contain the item string of "rhoma-irama" (such as name directory, page name) will still be stored in *.Rda file. If the link is not related to item string, the link will be deleted. On other hands, URL address filter is identifying the structure of link based on full of URL address, for example, if there is URL address filter "https://liriklaguindonesia.net/rhoma-irama", then all of the links which they contain its URL address will be kept. The Figure 8 is illustrated 18 data of source and link which are represented in URL, they have been filtered based on direct string filter and URL address filter. There are five links which is not related to Rhoma Irama's song lyric. Each link has a source to make easier identify the page location of each song title.

3.2 Metadata Extraction

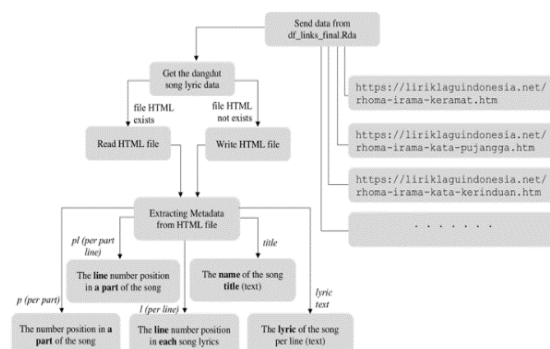


Figure 9. The metadata extraction to get p, pl, l, title, and lyric text of dangdut song from data store *.Rda file

After all filtered links are stored in *.Rda file, it will be continued to metadata extraction process. Figure 9 illustrated about how to check the existing of HTML file (contain the song lyric page like is illustrated in Figure 10) that is stored on local storage or not. If the HTML file exists, then HTML file will only be read (using `read_html` function in R programming). If it doesn't exist, then HTML file will be created using `write_html` function in R programming). To extract the metadata from HTML file, it requires a way to read the line numbers of each song lyrics.

There are five columns that must be created and filled in each row of data, including:

- **p (part):** identify the number position in a part of the song. For example: part-1, part-2, part-3, ... until part-n
- **pl (part line):** identify the line number position in a part of the song. For example: line-1 of part-1, line-2 of part-1, line-1 of part-2, line-2 of part-2, ... until line-n of part-n.
- **l (line):** identify the line number position in each song lyrics. For example: line-1, line-2, ... until line-n.

- **the song titles (text):** identify the name of the song title. For example: *Keramat, Kata Pujangga, Kerinduan*.
- **the song lyrics (text):** identify the lyric of the song per line. For example: *hai manusia hormati ibumu* (one line of dangdut song lyric with entitled *Keramat*).

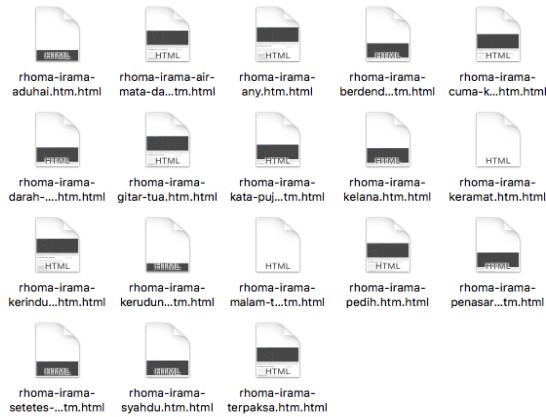


Figure 10. HTML file of Rhoma Irama's dangdut song lyric

All metadata will be stored in *.Rda file, Figure 11 illustrated there are 435 entries of lyric extraction with the number position detail of *p*, *pl*, and *l*. To identify *p* number, an HTML file that has a blank line is represented by `
` will be counted as a new part of the song. Therefore, when there is a blank line after part-1 it will be labelled with part-2. To identify *pl* number, in a part of the song consists of several lines of song lyrics, it will be counted during there are no new blank lines that will be grouped according to the current *p* number. Therefore, when there is a new blank line after part-1, each line in part-1 will be labelled line-1 of part-1, and line-2 of part-1. To identify *l*, all lines of song lyrics are counted as sequences begin line-1 to line-435.

source	title	p	pl	l	text
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	1	1	1	hai manusia hormati ibumu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	1	2	2	yang melahirkan dan membesarkanmu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	2	1	3	darah dagingmu dari air susunya
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	2	2	4	jika ragamu dari kasih sayangnya
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	2	3	5	dalah manusia satu-satunya
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	3	1	7	doa IbuMu dikabulkan tuhan
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	3	2	8	dan kutukannya jadi kenyataan
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	3	3	9	ridha lah karena ridhonya
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	3	4	10	murka lah karena murkanya
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	4	1	11	bila kau sayang pada kasihmu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	4	2	12	lebih sayang pada IbuMu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	4	3	13	bila kau patuh pada ragamu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	4	4	14	lebih patuh pada IbuMu
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	5	1	15	bukannya gunung tempat kau meminta
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	5	2	16	bukan lautan tempat kau memuja
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	5	3	17	bukan pula dilun tempat kau menghiba
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	6	1	19	tidak keramat yang ampuh di dunia
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	6	2	20	selain dari doa IbuMu juga
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	7	1	21	doa IbuMu dikabulkan tuhan
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	7	2	22	dan kutukannya jadi kenyataan
https://ir.kagondoesia.net/rhoma-irama-keramat.htm	Keramat	7	3	23	ridha lah karena ridhonya

Figure 11. The metadata extraction of dangdut song lyric is stored in *.Rda file.

To make an easier understanding, the given illustration in Figure 12 about how Rhoma Irama's dangdut song entitled "Keramat" is extracted its metadata:

- Part-1
 - Part-1 and Line-1 (lyric text): *hai manusia hormati ibumu*
 - Part-1 and Line-2 (lyric text): *yang melahirkan dan membesarkanmu*
- New blank line
- Part-2
 - Part-2 and Line-1 (lyric text): *darah dagingmu dari air susunya*
 - Part-2 and line-2 (lyric text): *jiwa ragamu dari kasih sayangnya*

Figure 12. The metadata representation of each song lyric (per line) in *.Rda file related to the structure of HTML file

3.3 Data Visualization

Data visualization plays an important role in how a text-based data that is difficult to read and detect information can be represented in visualizing patterns, trends, and correlations. Nowadays, data sets are not sufficiently processed through traditional calculations but require software that provides informative visualization. Datasets are easier to read in a chart than presented in a table. Data visualization has entered the scope of scientific writing (scientific visualization), raw data sets are processed as data visualization objects [9]. This paper also utilize data visualization to see analytic in a visual form of dangdut song lyric which consists of a bar chart, word cloud, *tf-idf*, and network graph.

Bar chart (horizontal line) in this paper is used to identify the number of frequencies of the most word lyrics on each dangdut song title. The data flow of bar chart visualization begin the process of importing metadata consist of *p*, *pl*, *l*, the title song, and lyric text from *.Rda file, and then data is grouped by title song. Each song counts the number of song part, the number of song line, and the total number of word in a song. The results of all frequency calculations for each dangdut song title will be sorted in the 15 top-ranking and visualized in the horizontal bar chart. In simple terms, a bar graph contains information on how many words are contained in one song.

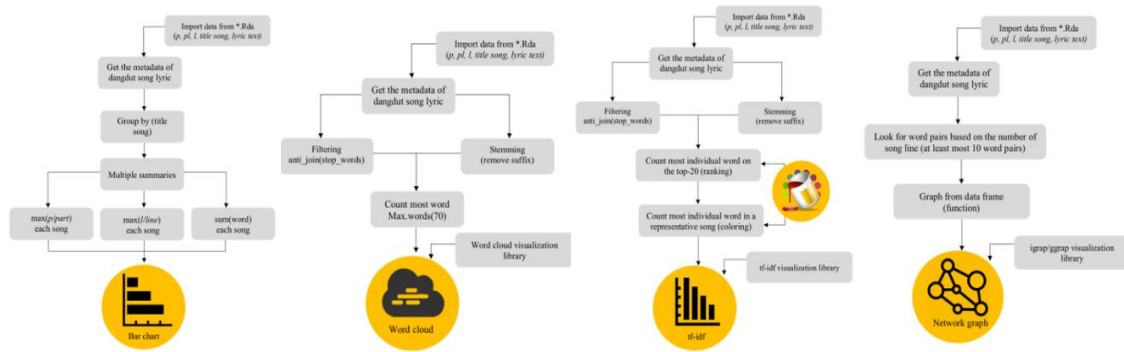


Figure 13. The data flow of visualization process to identify the most words frequency of the song on horizontal bar chart, word cloud, tf-idf, and network graph

The results of all frequency calculations for each dangdut song title will be sorted in the 15 top-ranking and visualized in the horizontal bar chart. In simple terms, a bar graph contains information on how many words are contained in one song.

Another form of visualization is **word cloud**, is used to identify the favourite word of the song with a keyword. The data flow of word cloud visualization also begin the process of importing metadata, and then lyric data is filtered based on stop words and carried out stemming based on the suffix. The form of dangdut song lyrics often ends with the words 'mu' (you), 'ku' (me), 'nya' (him/her), therefore stemming is necessary to get the basic word. For example, the lyric on the song title 'Keramat' (The sacred of mother) in line-1 '.....hai manusia hormati ibumu...' or in English is: extracted its metadata:

- 'hai': O
- 'manusia': people
- 'hormati': respect
- 'ibumu': your mather

The word 'ibumu' (your mother) must be removed the word 'mu' (your) to get basic word of 'ibu' (mother). Then, continued to count the most number of words based on the keyword that have at least 70 words in all of the songs. For example, using a keyword "love", "happy", and "longing".

tf-idf bar chart in this paper is used to identify the number of frequencies of the most individual words in a representative song with the coloring of the song title. The data flow of *tf-idf* visualization also begin the process of importing metadata, filtering, and stemming. Then, count the word in all of the dangdut songs. For example, the word lyric 'gembira' (happy) on the dangdut song X as much as Y with coloring is given to find out the difference of the song title in each word lyrics. The ranking presented on the *tf-idf* bar chart has at least 20 top-ranking song titles that have the most individual words.

Network graph is used to identify the most word pairs to visualize the word network. We know the songwriter usually write the song lyric with beautiful and interesting words. The word representation in the songs contains information about the song theme. The word choice in a song can describe what the meaning of a song is. Rhoma Irama as a dangdut songwriter also has a unique word lyrics choice for dangdut music lovers. He writes the songs with themes of romance, love, sadness, enthusiasm, and religious. Network graph can help to visualize the word pairs of the Rhoma Irama's dangdut song lyric. The two pairs of words in the network graph will look thick and thin in the word relationship line. If the line is thick then the relationship of the word is strong (most the frequency of word pairs), but if the line is thin then the relationship of the word is weak. Figure 13 illustrated the data flow of visualization in a horizontal bar chart, word cloud, *tf-idf*, and network graph.

4. Experiment Result

This section will be explained about collecting the dangdut song lyric data and the visualization result. The data collection process through the selection of Rhoma Irama's classic dangdut song that had been written in the 1970s. Indonesian people know that Rhoma Irama. Indonesians people know that Rhoma Irama is a legendary Indonesian singer who has many songs. Even he has a dream to make 1000 songs at the end of his career. Many albums have been created by him, but in this paper, only classic dangdut songs are chosen. While the visualization result needs to be interpreted into the knowledge creation to make it easy understand.

4.1 Collecting Dangdut Song Lyrics

We collected Rhoma irama's dangdut songs which were categorized as classic songs in the 1970s. Table 1 shows a list of his songs that are very popular in Indonesian. There are several song theme categories including romance

entitled ("Malam Terakhir", "Penasaran", "Cuma Kamu", "Air Mata Darah", "Ani", "Gitar Tua", "Kerinduan", "Syahdu", "Kerudung Putih", "Aduhai", "Kata Pujangga"), sadness entitled ("Kelana", "Terpaksa", "Pedih"), social entitled ("Keramat", "Setetes Air Hina", "Darah Muda"), and happiness entitled ("Berdendang").

Table 1. Dangdut song title list was created by Rhoma Rima and Group in 1970s era

No.	Dangdut Song Title	Vol. / Year
1.	Malam Terakhir	With Orkes Melayu Soneta / 1972
2.	Kelana	Soneta Volume Series 2 / 1974
3.	Penasaran	Soneta Volume Series 2 / 1974
4.	Cuma Kamu	Soneta Volume Series 4 / 1975
5.	Darah Muda	Soneta Volume Series 4 / 1975
6.	Air Mata Darah	Soneta Volume Series 3 / 1975
7.	Ani	Soneta Volume Series 6 / 1976
8.	Keramat	Soneta Volume Series 7 / 1977
9.	Gitar Tua	Movie Soundtrack of "Gitar Tua" / 1977
10.	Kerinduan	Movie Soundtrack of "Darah Muda" / 1977
11.	Berdendang	Movie Soundtrack of "Darah Muda" / 1977
12.	Terpaksa	Soneta Volume Series 9 / 1978
13.	Pedih	Movie Soundtrack of "Berkelana" / 1978
14.	Syahdu	Movie Soundtrack of "Berkelana" / 1978
15.	Kerudung Putih	Movie Soundtrack of "Perjuangan dan Doa" / 1980
16.	Setetes Air Hina	Soneta Volume Series 12 / 1981
17.	Aduhai	Movie Soundtrack of "Sebuah Pengorbanan" / 1982
18.	Kata Pujangga	Soneta Special Volume Series 1 / 1992

4.2 Data Visualization of Rhoma Irama's Dangdut Song

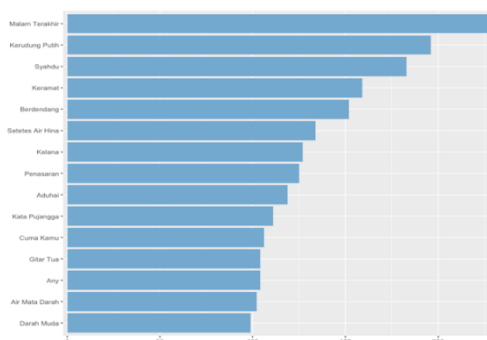


Figure 14. The horizontal bar chart of Rhoma Irama's Dangdut Song

Figure 14 shows the horizontal bar chart of 15 top-ranking on Rhoma Irama's dangdut songs with most of the words identified. The song entitled "Malam Terakhir" (The last night) has the most words compared to other songs, our data extraction in the *.Rda file analyzed there are 43 lines of this song lyrics. While the lowest ranking is the song entitled "Darah Muda" (The young blood spirit) in our data extraction analyzed there are only 16 lines of this song lyrics.



Figure 15. The word cloud of Rhoma Irama's Dangdut Song

Figure 15 shows the word cloud of Rhoma Irama's dangdut songs with the keyword "cinta" or "kasih" (love), "sayang" (dear, beloved), "hati" (heart), and "rindu" (longing). Our text mining data analyzed that the song entitled:

- "Syahdu" has the word "hati" (n=19, p=25.0%): romance
- "Kata Pujangga" has the word "cinta" (n=11, p=21.2%): romance
- "Ani" has the word "rindu" (n=6, p=13%): romance
- "Terpaksa" has the word "hati" (n=7, p=12.7%): sadness
- "Kerinduan" has the word "hati" (n=6, p=11.1%): romance
- "Kerinduan" has the word "rindu" (n=6, p=11.1%): romance
- "Air Mata Darah" has the word "hati" (n=4, p=8.89%): romance

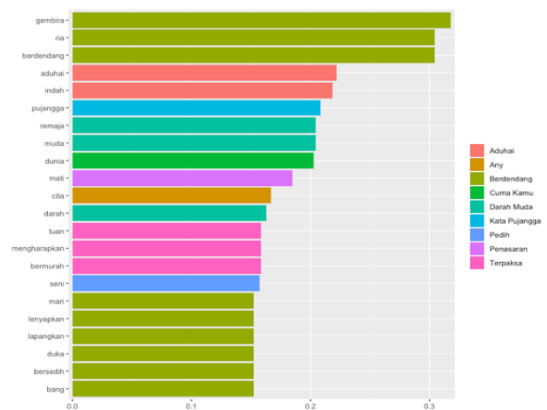


Figure 16. The td-idf bar chart of Rhoma Irama's Dangdut Song

Figure 16 shows the tf-idf (term frequency-inverse document frequency) bar chart of Rhoma Irama's dangdut songs, tf-idf is a measure of how important a word is to a document in a group of the corpus. Our text mining analyzed there are 20 top-ranking of important word including the word "gembira" (happy), "ria" (cheerful), "berdendang" (singing), and others.

Figure 17 shows the network graph, our text mining analyzed the word pairs between the word "hati" (heart) and "cinta" (love) has the thick line that represented the strong relationship of both. The strong relationship also was followed by the word "sayang" (dear, beloved) and "rindu" (longing).

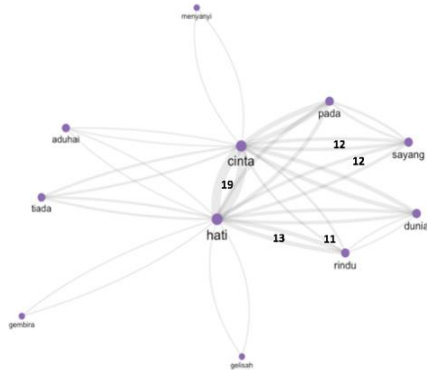


Figure 17. The network graph of Rhoma Irama's Dangdut Song

5. Conclusion

This paper proposed to explore Rhoma Irama's dangdut song lyric in 1970s era. Indonesian people have known the song themes that was written in his era was about love. We explored the relationship between the love song themes with the selected words through text mining. We discovered that there are four keywords that often was used by Rhoma Irama in writing their song including "hati" (heart), "cinta" (love), "sayang" (dear, beloved), and "rindu" (longing). The word cloud proved to analyze that they are the most words, while the network graph is increasingly proof that the strong relationship between word pairs of that all especially between the word "hati" (heart) and "cinta" (love). In the future works, it is necessary to explore the selected words and the relationship between word pairs of Rhoma Irama's dangdut song theme collection in the following year (1980, 1990, 2000, 2010 and so on). We hope from this research, the purity of dangdut songs will continue to be well known in the modern era by following the selected words as written by Rhoma Irama.

References

- [1] Tabloid Dangdut, No.4, Tahun 1, Juni 1995.
- [2] Aris S., "Dangdut: Budaya Musik Etnik Nusantara III", Dangdut Article, pp. 3.
- [3] Andrew N. Weintraub., "Dangdut Soul: Who are 'the People' in Indonesian Popular Music?", Asian Journal of Communication, Vol. 16, No. 4, pp. 411-431, 2006.
- [4] Frederick, W., "Rhoma Irama and The Dangdut Style: Aspects of Contemporary Indonesian Popular Culture", Southeast Asia Program Publications at Cornell University, No.34, pp.102-130, 1982.
- [5] Andrew N. Weintraub., "Dangdut Stories: A Social and Musical History of Indonesia's Most Popular Music", Oxford University Press, 1 edition, pp.1-272, 2010.
- [6] Daniel S., et al., "Genre Complexes in Popular Music", PLoS ONE Journal, Vol.11, No.5, pp.1-23, 2016.
- [7] Braja Gopal Patra., et al., "Mood Classification of Hindi Songs based on Lyrics", International Conference on Natural Language Processing (ICON), pp.261-267, 2015.
- [8] Chutimet Srinilta., et al., "Lyric-based Sentiment Polarity Classification of Thai Songs", International Multi Conference of Engineers and Computer Scientists (IMECS), Vol.1, 2017.
- [9] Matthew., et al., "Data Visualization", International Journal of Engineering Research and Advanced Technology (IJERAT), Vol.2, Issue.12, pp. 11-16, 2016.