

# MLDB Get Lyrics

Daniel Huang

21 September, 2021

Code to retrieve lyrics and other information from mldb.org search with artist and title.

```
library(stringr)

## Warning: package 'stringr' was built under R version 4.0.5

library(readr)

## Registered S3 methods overwritten by 'tibble':
##   method      from
##   format.tbl  pillar
##   print.tbl   pillar

library(magrittr)
library(rvest)

## Loading required package: xml2

##
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':
##
##   guess_encoding

get_lyrics <- function(title, artist) {
  #Code to return artist, album(s), and lyrics from MLDB.org database corresponding
#to a certain title and artist
#Uses stringr, readr, magrittr, rvest packages

  #Check inputs are both strings
  if (typeof(title) != "character" ||
      typeof(artist) != "character") {
    stop("Please enter character strings")
  }

  #Change artist and song name into correct form
  artist_mod <- str_to_lower(str_replace_all(artist, " ", "-"))
  title_mod <- str_replace_all(title, " ", "+")

  #Make search URL
  search_URL <-
    paste0("http://mldb.org/search?mq=",
           title_mod,
           "&si=2&mm=0&ob=1")
}
```

```

#Check if URL brought you to results page or directly to song page
check <-
  search_URL %>% read_html() %>% html_nodes("title") %>% html_text()
direct <- TRUE
if (str_detect(check, "Advanced search")) {
  direct <- FALSE
  #Check if there are any results, if none return NULL
  check2 <- search_URL %>% read_html() %>% html_nodes("#thelist")
  if (length(check2) < 1) {
    return (NULL)
  }
}

if (direct) {
  #If it goes directly then just get the info from the page
  results <- search_song_page(search_URL)
} else {
  #Make API call, gets table of hyperlinks with artists and songs
  #Table format is artist, song, artist, song, etc, so
  #we check each odd valued entry for the artist, if matches then we take
  #the next entry which is the correct song.

  start <- 0
  search_URL <- paste0(search_URL, "&from=0")

  #Boolean for making sure artist is found
  found <- FALSE

  #30 entries per page if full, so if full and we don't find the song we check next page
  while (start %% 30 == 0) {
    search_results <-
      search_URL %>% read_html() %>% html_nodes("#thelist") %>% html_nodes("a") %>% html_attr("href")

    num_results <- length(search_results) / 2

    for (i in 1:num_results) {
      j <- 2 * i - 1
      if (str_detect(search_results[j], artist_mod)) {
        song_URL <-
          paste0("http://mldb.org/", search_results[j + 1])
        found <- TRUE
      }
    }

    if (found) {
      break
    } else {
      start <- start + num_results
      search_URL <-
        str_replace(search_URL,
          paste0("&from=", start - num_results),

```

```

        paste0("&from=", start))
    }
}

if (found) {
  results <- search_song_page(song_URL)
} else {
  return (NULL)
}
return(results)
}
}

#Once we find the page of the song, searches it for the information we want (artist, lyrics, album(s))
search_song_page <- function(song_URL) {
  #Find artist + album
  artist_and_albums <-
    song_URL %>% read_html() %>% html_nodes("#thelist") %>% .[[1]] %>% html_table()

  #Find lyrics
  lyrics <-
    song_URL %>% read_html() %>% html_nodes("p.songtext") %>% html_text()

  #Results will be in a 3 element list. Order is artist, album(s), then lyrics.
  results <-
    list("artist" = artist_and_albums[1, 2],
         "album(s)" = artist_and_albums[2, 2],
         "lyrics" = lyrics)

  return(results)
}

#Testing cases
#Case 1: Standard case, go to search page then find song
print(get_lyrics("Dancing in the dark", "Bruce Springsteen"))

## $artist
## [1] "Bruce Springsteen"
##
## $`album(s)`
## [1] "Born in the U.S.A.Greatest Hits"
##
## $lyrics
## [1] "I get up in the evening, and I ain't got nothing to say\nI come home in the morning, I go to bed"

#Case 2: Goes directly to song page
print(get_lyrics("leaving las vegas", "sheryl crow"))

## $artist
## [1] "Sheryl Crow"
##
## $`album(s)`
## [1] "Tuesday Night Music ClubC'mon C'mon"
##

```

```

## $lyrics
## [1] "Life springs eternal\nOn a gaudy neon street\nNot that I care at all\nI spent the best part of m

#Case 3: Song does not exist
print(get_lyrics("adsfsadfdsaf", "hi"))

## NULL

#Case 4: Song exists, but artist is wrong
print(get_lyrics("Kiss Kiss", "joe biden"))

## NULL

#Case 5: Not on first page
print(get_lyrics("In the dark", "Peter Cetera"))

## $artist
## [1] "Peter Cetera"
##
## $`album(s)`
## [1] "One More Story"
##
## $lyrics
## [1] "Moment by moment, day after day\nTrying your best to forget\nIt's all behind you, it's part of

```