

VerySimpleRecommendationSystem

```
library(utils)
library(jsonlite)

#inspired from Handbook 1 Big Data Engineering Data Science Council of America
create_artist_query_url_lfm <- function(artist.name) {
  prefix <- "http://ws.audioscrobbler.com/2.0/?method=artist.getoptags&artist="
  postfix <- "&api_key=5075646ac58a83c2e351e6ea5ed497a4&format=json"
  encoded.artist <- URLencode(artist.name)
  return(paste0(prefix,encoded.artist, postfix))
}

get_tag_vector_lfm <- function(artist) {
  artist_url <- create_artist_query_url_lfm(artist)
  json <- fromJSON(artist_url)
  return(json$toptags$tag$name)
}

jaccard_index <- function(set1, set2) {
  length(intersect(set1, set2))/length(union(set1, set2))
}

similarity_matrix <- function(artist_list, similarity_fn) {
  n <- length(artist_list)
  #init n by n matrix with zeros
  sim_matrix <- matrix(0, ncol = n, nrow = n)

  #name rows and cols
  rownames(sim_matrix) <- names(artist_list)
  colnames(sim_matrix) <- names(artist_list)

  #for each row in the matrix
  for(i in 1:nrow(sim_matrix)) {
    #and each column
    for(j in 1:ncol(sim_matrix)) {
      #calculate similarity
      the_index <- similarity_fn(artist_list[[i]], artist_list[[j]])
      sim_matrix[i,j] <- round(the_index, 2)
    }
  }
  return(sim_matrix)
}

create_artist_query_url_lfm("Depeche Mode")

## [1] "http://ws.audioscrobbler.com/2.0/?method=artist.getoptags&artist=Depeche%20Mode&api_key=5075646ac58a83c2e351e6ea5ed497a4&format=json"
```

```

fromJSON(create_artist_query_url_lfm("Rammstein"))

## $toptags
## $toptags$tag
##   count           name          url
## 1    100 industrial metal https://www.last.fm/tag/industrial+metal
## 2     71            metal https://www.last.fm/tag/metal
## 3     66      industrial https://www.last.fm/tag/industrial
## 4     58        german https://www.last.fm/tag/german
## 5     37         rock https://www.last.fm/tag/rock
## 6     24      seen live https://www.last.fm/tag/seen+live
## 7     12    hard rock https://www.last.fm/tag/hard+rock
## 8      9 heavy metal https://www.last.fm/tag/heavy+metal
## 9      8      Rammstein https://www.last.fm/tag/Rammstein
## 10     6 neue deutsche härte https://www.last.fm/tag/neue+deutsche+h%C3%A4rte
## 11     6 industrial rock https://www.last.fm/tag/industrial+rock
## 12     4      deutsch https://www.last.fm/tag/deutsch
## 13     4      tanz metal https://www.last.fm/tag/tanz+metal
## 14     4 Neue Deutsche Haerte https://www.last.fm/tag/Neue+Deutsche+Haerte
## 15     4      alternative https://www.last.fm/tag/alternative
## 16     3      electronic https://www.last.fm/tag/electronic
## 17     3          NDH https://www.last.fm/tag/NDH
## 18     3       germany https://www.last.fm/tag/germany
## 19     2      dance metal https://www.last.fm/tag/dance+metal
## 20     2      alternative metal https://www.last.fm/tag/alternative+metal
## 21     2      German metal https://www.last.fm/tag/German+metal
## 22     2      tanz-metall https://www.last.fm/tag/tanz-metall
## 23     2       Nu Metal https://www.last.fm/tag/Nu+Metal
## 24     2      alternative rock https://www.last.fm/tag/alternative+rock
## 25     2          90s https://www.last.fm/tag/90s
## 26     2      Gothic Metal https://www.last.fm/tag/Gothic+Metal
## 27     1      german rock https://www.last.fm/tag/german+rock
## 28     1      male vocalists https://www.last.fm/tag/male+vocalists
## 29     1      Tanzmetal https://www.last.fm/tag/Tanzmetal
##
## $toptags$`@attr`$artist
## [1] "Rammstein"

get_tag_vector_lfm("Rammstein")

## [1] "industrial metal"      "metal"          "industrial"
## [4] "german"                 "rock"           "seen live"
## [7] "hard rock"               "heavy metal"    "Rammstein"
## [10] "neue deutsche härte"   "industrial rock" "deutsch"
## [13] "tanz metal"              "Neue Deutsche Haerte" "alternative"
## [16] "electronic"              "NDH"             "germany"
## [19] "dance metal"              "alternative metal" "German metal"
## [22] "tanz-metall"             "Nu Metal"        "alternative rock"
## [25] "90s"                     "Gothic Metal"   "german rock"
## [28] "male vocalists"          "Tanzmetal"       ""

artists <- list("Eminem", "Rammstein", "Avril Lavigne", "The Offspring", "Slipknot", "Metallica")
artists.tags <- lapply(artists, get_tag_vector_lfm)
names(artists.tags) <- artists

```

```

print(artists.tags)

## $Eminem
## [1] "rap"           "Hip-Hop"        "Eminem"
## [4] "hip hop"       "pop"            "american"
## [7] "detroit"        "seen live"       "hiphop"
## [10] "Gangsta Rap"   "male vocalists" "slim shady"
## [13] "rock"          "alternative"    "hardcore rap"
## [16] "90s"           "00s"            "USA"
## [19] "singer-songwriter" "Shady"          "Horrorcore"
## [22] "d12"          "emo"            "Aftermath"
## [25] "dance"         "Michigan"       "white rap"
##
## $Rammstein
## [1] "industrial metal" "metal"          "industrial"
## [4] "german"          "rock"           "seen live"
## [7] "hard rock"        "heavy metal"     "Rammstein"
## [10] "neue deutsche härte" "industrial rock" "deutsch"
## [13] "tanz metal"      "Neue Deutsche Haerte" "alternative"
## [16] "electronic"      "NDH"            "germany"
## [19] "dance metal"    "alternative metal" "German metal"
## [22] "tanz-metall"    "Nu Metal"        "alternative rock"
## [25] "90s"             "Gothic Metal"    "german rock"
## [28] "male vocalists" "Tanzmetal"       ""
##
## $`Avril Lavigne`:
## [1] "pop"            "pop rock"
## [3] "rock"           "female vocalists"
## [5] "Canadian"       "punk"
## [7] "pop punk"       "alternative"
## [9] "avril lavigne" "seen live"
## [11] "punk rock"      "female"
## [13] "alternative rock" "singer-songwriter"
## [15] "female vocalist" "Pop-Rock"
## [17] "Avril"          "00s"
## [19] "emo"            "canada"
## [21] "teen pop"      "indie"
## [23] "Pop-punk"       "acoustic"
## [25] "Officially Shit" "dance"
## [27] "sexy"           "guilty pleasure"
## [29] "shit"           "favorites"
## [31] "Brutal Death Metal" "acoustic pop"
## [33] "Mellow"          "Crap"
## [35] "Sandy Leah"     "power pop"
## [37] "female artists" "cute"
## [39] "black metal"    "Love"
## [41] "female vocals"  "english"
## [43] "folk"            "catchy"
## [45] "female vocal"   "nsbm"
## [47] "post-grunge"    "soft rock"
## [49] "melancholy"     "sandy e junior"
## [51] "diva"            "hot"
## [53] "punk pop"        "electronic"
## [55] "Guilty Pleasures" "RAC"

```

```

## [57] "poprock"                                "Acoustic Rock"
## [59] "national socialist black metal"          "depressive black metal"
## [61] "Orthodox Black Metal"                   "Rock Against Communism"
## [63] "noise"                                    "Awesome"
## [65] "atmospheric black metal"                "hatecore"
## [67] "true metal"                             "Post-Black Metal"
## [69] "blackgaze"                             "Blackwave"
##
## $`The Offspring`
## [1] "punk rock"      "punk"           "rock"           "alternative"
## [5] "alternative rock" "seen live"       "pop punk"        "american"
## [9] "hard rock"       "90s"            "The Offspring"  "skate punk"
## [13] "metal"          "California"     "Pop-punk"       "Punk-Rock"
## [17] "USA"            "hardcore"        "offspring"      "indie"
## [21] "Grunge"         "punkrock"       "pop"            ""
##
## $Slipknot
## [1] "Nu Metal"       "metal"          "heavy metal"
## [4] "Nu-metal"        "seen live"      "rock"
## [7] "alternative metal" "alternative"   "Slipknot"
## [10] "hard rock"       "metalcore"      "hardcore"
## [13] "death metal"    "american"      "thrash metal"
## [16] "mallcore"        "alternative rock" "new metal"
## [19] "industrial metal" "NOT METAL"    "groove metal"
## [22] "shit"           "gay"            "rapcore"
## [25] "emo"             "USA"            "industrial"
## [28] "crossover"       "iowa"           "trash metal"
## [31] "punk"             "pop"            "90s"
## [34] "Crap"            "corey taylor"   ""
##
## $Metallica
## [1] "thrash metal"   "metal"          "heavy metal"
## [4] "hard rock"       "rock"           "seen live"
## [7] "metallica"       "speed metal"    "classic rock"
## [10] "american"       "thrash"         "80s"
## [13] "trash metal"    "90s"            "emo"
## [16] "USA"             "alternative"   "classic metal"
## [19] "Progressive metal" "alternative rock" "heavy"
## [22] "legends"        "Power metal"    ""
jaccard_index(artists.tags[["Rammstein"]], artists.tags[["Avril Lavigne"]])

## [1] 0
similarity.artists <- similarity_matrix(artists.tags, jaccard_index)
print(similarity.artists)

##          Eminem Rammstein Avril Lavigne The Offspring Slipknot Metallica
## Eminem      1.00      0.10      0.09      0.16      0.15      0.16
## Rammstein    0.10      1.00      0.05      0.16      0.23      0.18
## Avril Lavigne  0.09      0.05      1.00      0.12      0.09      0.06
## The Offspring 0.16      0.16      0.12      1.00      0.26      0.24
## Slipknot      0.15      0.23      0.09      0.26      1.00      0.29
## Metallica     0.16      0.18      0.06      0.24      0.29      1.00

```

```
#print an order for recommendation for Rammstein
similarity.artists[order(similarity.artists[,2], decreasing = TRUE), 2]

##      Rammstein      Slipknot      Metallica The Offspring      Eminem
##      1.00          0.23          0.18          0.16          0.10
## Avril Lavigne
##          0.05
```