**RETO 3**

**Estructuras de Datos y Algoritmos**

***Estudiante A:*** *Daniel Andrés Bernal Cáceres 202020706*

***Estudiante B:*** *Juan Martín Santos Ayala 202013610*

***Correos:*** *da.bernalc1@uniandes.edu.co*

*j.santosa* *@uniandes.edu.co*

1. **Preparación del ambiente de trabajo:**

Para la preparación del ambiente de trabajo, creamos los archivos correspondientes al modelo “MVC” para independizar los cambios hechos en código y que no se vieran reflejados en el producto entregado al usuario.

* 1. **Carga de archivos**

Para cargar los archivos primero creamos un analizador para guardar la información de un archivo de forma organizada. Tal organización fue:

def newAnalyzer():

    analyzer = {'event': None,

                'tracks': None,

                'artists': None,

                'instrumentalness': None,

                'acousticness': None,

                'liveness':  None,

                'speechiness': None,

                'energy': None,

                'danceability': None,

                'valence': None,

                'tempo': None,

                'hashtag': None

                }

Creamos una lista vacía con todas las pistas cargadas para facilitar la consulta del tamaño total de datos cargados del archivo CSV.

    analyzer['event'] = lt.newList('ARRAY\_LIST')

Se crean índices (Maps) por los siguientes criterios: ‘track\_id’

    analyzer['tracks'] = om.newMap(omaptype='RBT')

    analyzer['artists'] = om.newMap(omaptype='RBT')

    analyzer['instrumentalness'] = om.newMap(omaptype='RBT')

    analyzer['acousticness'] = om.newMap(omaptype='RBT')

    analyzer['liveness'] = om.newMap(omaptype='RBT')

    analyzer['speechiness'] = om.newMap(omaptype='RBT')

    analyzer['energy'] = om.newMap(omaptype='RBT')

    analyzer['danceability'] = om.newMap(omaptype='RBT')

    analyzer['valence'] = om.newMap(omaptype='RBT')

    analyzer['tempo'] = om.newMap(omaptype='RBT')

    analyzer['hashtag'] = mp.newMap(numelements=17,

                                    maptype='CHAINING',

                                    loadfactor=0.5)

Después de crear el catálogo lo que hacemos es agregar toda la información del archivo en una lista de tipo ARRAY\_LIST, esta la usamos para conocer el size que lo necesitamos en varias funciones.

def addVideoToCat(catalog, video):

    lt.addLast(catalog['videos'], video)

Después de crear el catálogo lo que hacemos es adicionar el track\_id de cada evento al map. Si el track\_id ya está en el árbol, se adiciona a su lista respectiva y se actualiza el index. Si no se encuentra creado un nodo para ese id en el árbol, se crea y se actualiza el indeice del id de las pistas.

Actualiza el track\_id de cada evento y se adiciona al map.

def updateIdIndex(maps, event):

    eventId = event['track\_id']

    entry = om.get(maps, eventId)

    if entry is None:

        datantry = newDataEntry(event)

        om.put(maps, eventId, datantry)

    else:

        datantry = me.getValue(entry)

    addEntry(datantry, event)

    return maps

Actualiza el artist\_id de cada evento y se adiciona al map.

def updateartistsIndex(maps, event):

    eventId = event['artist\_id']

    entry = om.get(maps, eventId)

    if entry is None:

        datantry = newDataEntry(event)

        om.put(maps, eventId, datantry)

    else:

        datantry = me.getValue(entry)

    addEntry(datantry, event)

    return maps

Después de crear el catálogo lo que hacemos es actualizar los arboles de cada una de las características del contenido. Así mismo, agregamos una lista con todos los eventos de escucha al árbol dependiendo del valor que estos tienen en esa característica.

def updateContCara(maps, event):

    caracteristics = ['instrumentalness', 'acousticness',

                      'liveness', 'speechiness', 'energy',

                      'danceability', 'valence', 'tempo']

    for i in caracteristics:

        entry = om.get(maps[i], float(event[i]))

        if entry is None:

            datantry = newDataEntry(event)

            om.put(maps[i], float(event[i]), datantry)

        else:

            datantry = me.getValue(entry)

        addEntry(datantry, event)

    return maps

Después añadimos un evento a la lista que es valor en cada uno de los mapas del analyzer.

def newDataEntry(event):

    entry = {'lstevent': None}

    entry['lstevent'] = lt.newList('ARRAY\_LIST')

    return entry

Después creamos una entrada en el índice por track\_id, es decir en el árbol binario.

def newDataEntry(event):

    entry = {'lstevent': None}

    entry['lstevent'] = lt.newList('ARRAY\_LIST')

    return entry

* 1. **Mostrar la información al usuario**

Por último, antes de comenzar a realizar los requerimientos, imprimimos la información al usuario. Acá fue donde nos dimos cuenta de que era mejor tener todos los videos en una lista en específico. Esto nos serviría para conocer el size del archivo a leer y en primera medida todos los datos del csv.

**La complejidad de nuestro catálogo es de O(n logn).**

1. **Requerimiento 1**

El requerimiento 1 nos exigía encontrar cuántas reproducciones se tienen en el sistema de recomendación basado en una característica de contenido y con un rango determinado.

def caracterizeReproductions(maps, characteristic, keylo, keyhi):

    exists = characteristic in maps

    if exists is True:

        events = om.values(maps[characteristic], keylo, keyhi)

*# creamos una tabla de has para guardar a los artistas únicos*

        artistsMap = mp.newMap(34500,

                               maptype='PROBING',

                               loadfactor=0.5)

        eventsq = 0

        iterator = ite.newIterator(events)

        while ite.hasNext(iterator):

*# este elemento tiene la lista con los eventos*

            eventlist = ite.next(iterator)

*# Se suman la cantidad de eventos*

            eventsq += lt.size(eventlist['lstevent'])

            newiterator = ite.newIterator(eventlist['lstevent'])

            while ite.hasNext(newiterator):

*# Este ya es el evento con sus caracteristicas*

                event = ite.next(newiterator)

*# Se agrega a una tabla de hash los id's de los artistas*

                mp.put(artistsMap, event['artist\_id'], 0)

*# Es la cantidad de artistas unicos del requerimiento*

        artistsq = mp.size(artistsMap)

        return eventsq, artistsq

    else:

        return 0

Al final de esta función obtenemos la cantidad de eventos que hay y la cantidad de artistas únicos, y en el view sólo imprimimos los que nos pide el usuario.

**La complejidad del requerimiento 1 es de O(2n logn).**

1. **Requerimiento 2 (Estudiante A)**

El requerimiento 2 nos exigía encontrar las pistas en el sistema de recomendación que pueden utilizarse en una fiesta.

def partyMusic(maps, keylo1, keyhi1, keylo2, keyhi2):

    easylist = studyMap(maps, keylo1, keyhi1, keylo2, keyhi2,

                        'energy', 'danceability')

    if easylist is None:

        return 0

    else:

        tracksMap = mp.newMap(34500,

                              maptype='PROBING',

                              loadfactor=0.5)

        iterator = ite.newIterator(easylist)

        while ite.hasNext(iterator):

            events = ite.next(iterator)

            mp.put(tracksMap, events['track\_id'], events)

        size = lt.size(tracksMap)

        if size > 5:

            tracklist = 0

        else:

            lista = mp.valueSet(tracksMap)

            tracklist = get5artists(tracksMap, lista, size)

        return size, tracklist

Busca dentro del mapa deseado, en un rango especificado por el usuario, los eventos con una pista única que cumplen el rango especificado.

def studyMap(maps, keylo1, keyhi1, keylo2, keyhi2, caract1, caract2):

*# Aca se organizaran los eventos por segunda vez*

    eventsmap = om.newMap(omaptype='RBT')

*# Es la lista con todos los eventos de instrumentalness*

    eventslist = om.values(maps[caract1], keylo1, keyhi1)

    if lt.isEmpty(eventslist):

        return None

    else:

        iterator = ite.newIterator(eventslist)

        while ite.hasNext(iterator):

            events = ite.next(iterator)

            newiterator = ite.newIterator(events['lstevent'])

            while ite.hasNext(newiterator):

                event = ite.next(newiterator)

                om.put(eventsmap, float(event[caract2]), event)

        return om.values(eventsmap, keylo2, keyhi2)

Función que me retorna 5 valores al azar de una lista de listas pasada por parámetro

def get5artists(map, lista, size):

    finallist = lt.newList("ARRAY\_LIST")

    eventlist = random.sample(range(size), 5)

    for i in eventlist:

        keya = lt.getElement(lista, i)

        lt.addLast(finallist, keya)

    return finallist

Al final de esta función obtenemos una tupla, donde en la primera posición se encuentra la llave con la información del video más tendencia y en la segunda posición se encuentra una lista con la cantidad de veces que el video fue tendencia.

**La complejidad de nuestro requerimiento 2 es de O(n).**

1. **Requerimiento 3 (Estudiante B)**

El requerimiento 3 nos exigía encontrar las pistas en el sistema de recomendación que pueden utilizarse en una fiesta.

def partyMusic(maps, keylo1, keyhi1, keylo2, keyhi2):

    easylist = studyMap(maps, keylo1, keyhi1, keylo2, keyhi2,

                        'energy', 'danceability')

    if easylist is None:

        return 0

    else:

        tracksMap = mp.newMap(34500,

                              maptype='PROBING',

                              loadfactor=0.5)

        iterator = ite.newIterator(easylist)

        while ite.hasNext(iterator):

            events = ite.next(iterator)

            mp.put(tracksMap, events['track\_id'], events)

        size = lt.size(tracksMap)

        if size > 5:

            tracklist = 0

        else:

            lista = mp.valueSet(tracksMap)

            tracklist = get5artists(tracksMap, lista, size)

        return size, tracklist

Busca dentro del mapa deseado, en un rango especificado por el usuario, los eventos con una pista única que cumplen el rango especificado.

def studyMap(maps, keylo1, keyhi1, keylo2, keyhi2, caract1, caract2):

*# Aca se organizaran los eventos por segunda vez*

    eventsmap = om.newMap(omaptype='RBT')

*# Es la lista con todos los eventos de instrumentalness*

    eventslist = om.values(maps[caract1], keylo1, keyhi1)

    if lt.isEmpty(eventslist):

        return None

    else:

        iterator = ite.newIterator(eventslist)

        while ite.hasNext(iterator):

            events = ite.next(iterator)

            newiterator = ite.newIterator(events['lstevent'])

            while ite.hasNext(newiterator):

                event = ite.next(newiterator)

                om.put(eventsmap, float(event[caract2]), event)

        return om.values(eventsmap, keylo2, keyhi2)

Función que me retorna 5 valores al azar de una lista de listas pasada por parámetro

def get5artists(map, lista, size):

    finallist = lt.newList("ARRAY\_LIST")

    eventlist = random.sample(range(size), 5)

    for i in eventlist:

        keya = lt.getElement(lista, i)

        lt.addLast(finallist, keya)

    return finallist

Al final de esta función obtenemos una tupla, donde en la primera posición se encuentra la llave con la información del video más tendencia y en la segunda posición se encuentra una lista con la cantidad de veces que el video fue tendencia.

**La complejidad de nuestro requerimiento 3 es de O(n).**

1. **Requerimiento 4**

El requerimiento 4 nos exigía encontrar los videos con más likes en un país y con una etiqueta en específico.

def mostLikedVideosCountryTag(catalog, country, tag):

*# Obtenemos la pareja llave valor de la categoria*

    pair = mp.get(catalog['country'], country)

    if pair is None:

        newList = 1

    else:

*# Obtenemos la lista con los videos de ese pais*

        countryList = me.getValue(pair)

        finalList = lt.newList('ARRAY\_LIST')

        iterator = ite.newIterator(countryList['videos'])

        while ite.hasNext(iterator):

            info = ite.next(iterator)

*# Buscamos los tags que desea el usuario*

            if tag in info['tags']:

*# Verificamos que no hayan repeticiones de videos*

                lt.addLast(finalList, info)

*# Le hacemos un sort a la lista dependiendo de los likes*

        newList = sortVideos(finalList, cmpVideosByLikes)

        return newList

Al final de esta función obtenemos una lista, donde se encuentran los videos con más likes del país dado por parámetro y que tienen la etiqueta dada.

**La complejidad de nuestro requerimiento 4 es de O(n)**

* Comparar la complejidad de los requerimientos implementados en el Reto No. 1 con los implementados en este reto.

**Requerimiento 1:**

En nuestro reto 1 la complejidad era de O(n logn) y en el reto 2 es de O(n logn).

**Requerimiento 2:**

En nuestro reto 1 la complejidad era de O(2n^2) y en el reto 2 es de O(n).

**Requerimiento 3:**

En nuestro reto 1 la complejidad era de O(2n^2) y en el reto 2 es de O(n).

**Requerimiento 4:**

En nuestro reto 1 la complejidad era de O(2n) y en el reto 2 es de O(n).

Estas complejidades fueron calculadas teniendo en cuenta el peor de los casos.

Dicho esto, se puede concluir que los diferentes algoritmos utilizados para el reto 2 implementando los maps (tablas de Hash) son claramente más eficientes tanto en la búsqueda como también para agregar elementos, permitiéndonos tener búsquedas con complejidades logarítmicas. Conllevando, a que los tiempos fueran menores a los del reto 1.