Estructuras de Datos y Algoritmos

Jose Luis Tavera - 201821999 Juan Diego Yepes - 202022391

Reto II

1 Construcción del Catálogo

Para dar solución a los requerimientos, fue necesario replantear la organización del catálogo del reto 1, para organizar la información en mapas en vez de listas.

Funciones del View

Para empezar a inicializar el catálogo y cargar la información, el usuario debe digitar las siguientes opciones 1 y 2, las cuales empiezan las siguientes instrucciones:

```
if int(inputs[0]) == 1:
          x = controller.initCatalog()
      elif str(inputs[0]) == "2":
          print("\nCargando informacion de los archivos ....")
          print('_____
          answer = controller.loadData(x)
          print('Videos cargados: ' + str(controller.videosSize(x)))
          print("Tiempo [ms]: ", f"{answer[0]:.3f}", " || ",
                 "Memoria [kB]: ", f"{answer[1]:.3f}")
          print('\n')
11
          loaded_categories = controller.categoriesSize(x)
          print('Categorias cargadas: ' + str(loaded_categories))
          print_table(x)
14
          print('\n')
          loaded_countries = controller.countriesSize(x)
16
          print('Pa ses cargados: ' + str(controller.countriesSize(x)))
17
          print_list(x)
```

Estas funciones hacen referencia a las funciones en el controller que se exponen a continuación. A su vez, imprimen la información retorno de lo que se cargó en el catálogo; es decir cuántas y cuáles categorías se cargaron, cuantos y cuáles países, etc.

Funciones del Controller

Las funciones utilizadas en el controller se detallan a continuación.

```
def initCatalog():
    """

Llama la funcion de inicializacion del catalogo del modelo.
    """

catalog = model.newCatalog()
    return catalog
```

Esta función sirve de intermediaria con la función del model llamada newCatalog().

```
def loadData(catalog):
"""
Carga los datos de los archivos y cargar los datos en la
```

```
estructura de datos, a su vez mide el tiempo de ejecuci n
4
5
      delta\_time = -1.0
6
      delta_memory = -1.0
      tracemalloc.start()
9
10
      start_time = getTime()
      start_memory = getMemory()
11
      loadVideos(catalog)
13
      loadCategories(catalog)
14
15
      stop_memory = getMemory()
16
      stop_time = getTime()
17
      tracemalloc.stop()
18
19
20
      delta_time = stop_time - start_time
      delta_memory = deltaMemory(start_memory, stop_memory)
21
22
      return delta_time, delta_memory
```

Esta función carga los datos de los archivos y mide el tiempo de ejecución de las operaciones utilizando la librería time. Para cargar los datos, referencia las funciones loadVideos() y loadCategories(), las cuales se exponen a continuación:

```
def loadVideos(catalog):
3
      Carga los videos del archivo.
      videosfile = cf.data_dir + 'videos/videos-large.csv'
      input_file = csv.DictReader(open(videosfile, encoding='utf-8'))
      for video in input_file:
          model.addVideo(catalog, video)
  def loadCategories(catalog):
2
      Carga todas las categor as del archivo y los agrega a la lista de
3
      categorias
4
      categoriesfile = cf.data_dir + 'videos/category-id.csv'
      input_file = csv.DictReader(
          open(categoriesfile, encoding='utf-8'), delimiter='\t')
8
      for category in input_file:
9
          model.addCategory(catalog, category)
10
```

Estas funciones son las que leen el archivo .csv y van llamando a las otras funciones que agregan la información en el catálogo, las cuales se encuentran en el model.

Funciones del Model

Esta primera función inicializa el catálogo vacío, especificando cuáles índices van a ser mapas y cuáles van a ser listas. Una vez inicializado el catálogo, se agrega la información mediante las siguientes funciones:

Luego, las funciones de carga de datos, llamadas para cada video individual en el controller, se exponen a continuación:

```
def addVideo(catalog, video):
    """

La funcion de addVideo() adiciona los videos a una lista de videos,
    a su vez los adiciona a los mapas de 'category' y 'country'

"""

lt.addLast(catalog['videos'], video)
    addVideoOnMap(catalog, video['category_id'], video, 'category')
    addVideoOnMap(catalog, video['country'], video, 'country')
```

La función addVideo() no tiene retorno y se encarga de agregar individualmente, video por video, las funciones al catálogo. Primero, lo agrega a la lista de videos catalog['videos']; luego los agrega a cada uno de los mapas video['category_id'] y video['country']. Para hacer esto, se referencia a la función addVideoOnMap() la cual se muestra a continuación:

```
def addVideoOnMap(catalog, int_input, video, catalog_key):
      La funcion de addVideoOnMap() adiciona el video al mapa
      que se ha seleccionado.
      Args:
          catalog: Catalogo de videos
6
          int_input: Llave a analizar
          video: Video a a adir
          catalog_key: Especifica cu l catalogo
9
      selected_map = catalog[catalog_key]
11
      existkey = mp.contains(selected_map, int_input)
12
13
      if existkey:
          entry = mp.get(selected_map, int_input)
14
15
          value = me.getValue(entry)
16
17
          value = newSeparator(int_input, catalog_key)
          mp.put(selected_map, int_input, value)
18
      lt.addLast(value['videos'], video)
```

Esta función agrega el video individual al mapa seleccionado, que se define con el parámetro 'catalog_key'. Si la llave del video, que puede ser la categoría o el país, ya se encuentra en el mapa, lo que hace la función es agregarlo a esa llave expandiendo la lista de valores organizada en Array. Si no, crea una nueva lista y una nueva llave organizada en Array y agrega ese valor.

2 Primer Requerimiento

El equipo de análisis quiere conocer cuáles son los n videos con más views que son tendencia en un determinado país, dada una categoría específica. Para dar respuesta a este requerimiento el equipo de desarrollo debe recibir como entrada la siguiente información:

- category_name
- country
- Número de videos que quiere listar (n)

Y como respuesta debe presentar en consola la siguiente información:

- trending_date
- title
- channel_title
- publish_time
- views
- likes
- dislikes

Antes de analizar los ordenes de crecimiento del requerimiento, señalaremos las funciones y sus tareas dentro del modelo MVC. Por lo tanto, empezaremos por el view, seguido del controller y finalmente haremos el análisis del orden principalmente en las funciones del model.

Funciones del View

El view le pide al usuario ingresar el país y la categoría de referencia que servirán como parámetro para la función del controller que une con la función homónima

```
elif str(inputs[0]) == "3":
    pais = input("\nIngrese el pa s de referencia: ")

categoria = int(
    input('Ingrese el c digo la categor a de referencia: '))

n = input("Ingrese el n mero de videos que desea imprimir: ")

print("\nCargando ....")

result = controller.getVideosByCategoryAndCountry(x, categoria, pais)

printResults(result[1], int(n))
```

Funciones del Controller

Análogamente, el controller invoca la función homónima del model haciendo de puente entre el view y el model.

Funciones del Model

La función general del model, similar a la del reto 1, invoca tres funciones: dos para crear las sublistas y otro para sortear los videos.

```
def getVideosByCategoryAndCountry(catalog, category, country):
    """

La funci n de getVideosByCriteriaMap() filtra los videos por una
    llave y categor a espec ficas.
    """

sublist = getVideosByCriteriaMap(
    catalog, 'category', category).get('videos')

sublist2 = getVideosByCriteriaList(sublist, 'country', country)
return sortVideos(sublist2, lt.size(sublist2), 'cmpVideosByViews')
```

La función de Criteria Map filtra los vídeos por categoría usando el mapa del catálogo.

```
def getVideosByCriteriaMap(catalog, criteria, key):
    """

La funci n de getVideosByCriteriaMap() filtra los videos por un
    criterio espec fico dado un key. El cat logo debe ser un mapa.
    """

values = catalog[criteria]
entry = mp.get(values, str(key))
result = me.getValue(entry)
return result
```

La función de CriteriaList itera la lista para filtrar por el segundo criterio que corresponde al país.

```
def getVideosByCriteriaList(catalog, criteria, x):
      La funci n de getVideosByCriteriaList() filtra los videos por un
3
      criterio espec fico dado un x. El cat logo debe ser una lista.
5
      listaretorno = lt.newList("ARRAY_LIST")
6
      for element in lt.iterator(catalog):
8
          nombre_pais = element.get(criteria)
          if nombre_pais == x:
9
              lt.addLast(listaretorno, element)
10
   return listaretorno
12
```

Finalmente, sortVideos organiza la sublista final por merge sort para obtener los videos con más views por medio de la función de cmpVideosbyViews.

```
def sortVideos(catalog, size, cmp):
      La Funci n sortVideos() organiza los videos de acuerdo
3
      al par metro cmp.
4
      sub_list = lt.subList(catalog, 0, size)
6
      sub_list = sub_list.copy()
      start_time = time.process_time()
9
      if cmp == 'cmpVideosByViews':
10
          sorted_list = ms.sort(sub_list, cmpVideosByViews)
11
12
      if cmp == 'comparetitles':
          sorted_list = ms.sort(sub_list, comparetitles)
13
      if cmp == 'comparelikes':
14
          sorted_list = ms.sort(sub_list, comparelikes)
15
16
      stop_time = time.process_time()
17
      elapsed_time_mseg = (stop_time - start_time)*1000
18
return elapsed_time_mseg, sorted_list
```

3 Segundo Requerimiento (Juan Diego Yepes)

El equipo de análisis quiere conocer cuál es el video que más días ha sido trending para un país específico. Para dar respuesta a este requerimiento el equipo de desarrollo debe recibir como entrada la siguiente información:

• country

Y como respuesta debe presentar en consola la siguiente información:

- title
- channel_title
- country
- número de días

Antes de analizar los ordenes de crecimiento del requerimiento, señalaremos las funciones y sus tareas dentro del modelo MVC. Por lo tanto, empezaremos por el view, seguido del controller y finalmente haremos el análisis del orden principalmente en las funciones del model.

Funciones del View

El segundo requerimiento es invocado en el view cuando el usuario ingresa la opción número "4":

Para filtrar los videos, se invoca la función getVideosByCountry() del controller.

Funciones del Controller

Las funciones utilizadas en el controller se detallan a continuación.

Esta función a su vez invoca la función del model getVideosByCriteriaMap()

Funciones del Model

```
def getVideosByCriteriaMap(catalog, criteria, key):
    """

La funcion de getVideosByCriteriaMap() filtra los videos por un
    criterio espec fico dado un key. El cat logo debe ser un mapa.
    """

values = catalog[criteria]
entry = mp.get(values, str(key))
result = me.getValue(entry)
return result
```

La función de Criteria Map filtra los vídeos por categoría usando el mapa del catálogo.

```
def getMostTrendingDaysByID(videos):
      La funci n de getMostTrendingDaysByID() itera la lista ordenada y
3
      retorna el elemento que m s se repite.
5
6
      elemento = lt.firstElement(videos)
      mayor_titulo = None
      mayor = 0
8
      i = 0
9
10
11
      for video in lt.iterator(videos):
          if video['video_id'] == elemento['video_id']:
12
              i += 1
13
14
          else:
              if i > mayor:
15
                  mayor_titulo = elemento
mayor = i
16
17
              i = 1
18
19
              elemento = video
20
      if i > mayor:
21
22
          mayor_titulo = elemento
23
          mayor = i
     return (mayor_titulo, mayor)
24
```

4 Tercer Requerimiento (Jose Luis Tavera)

El equipo de análisis quiere conocer cuál es el video que más días ha sido trending para una categoría específica. Para dar respuesta a este requerimiento el equipo de desarrollo debe recibir como entrada la siguiente información:

• country

Y como respuesta debe presentar en consola la siguiente información:

- title
- channel_title
- country
- número de días

Antes de analizar los ordenes de crecimiento del requerimiento, señalaremos las funciones y sus tareas dentro del modelo MVC. Por lo tanto, empezaremos por el view, seguido del controller y finalmente haremos el análisis del orden principalmente en las funciones del model.

Funciones del View

Al seleccionar la opción cinco el usuario solo necesita ingresar la categoría de referencia para invocar dos funciones del controller: uno que filtra usando el mapa y otra que obtiene el vídeo que más veces aparece

Funciones del Controller

Análogamente, el controller cumple su función como puente entre el view y el model, invocando las funciones homónimas del model.

Funciones del Model

La función de getVideosByCriteriaMap usa los mapas creados en el catálogo para filtrar usando la categoría ingresada.

```
def getVideosByCriteriaMap(catalog, criteria, key):
    """

La funci n de getVideosByCriteriaMap() filtra los videos por un
    criterio espec fico dado un key. El cat logo debe ser un mapa.
    """

values = catalog[criteria]
    entry = mp.get(values, str(key))
    result = me.getValue(entry)
    return result
```

Inicialmente, la función necesita recibir la lista filtrada sorteada alfabéticamente para poder iterarla e ir guardando el video que más veces se repite.

```
def getMostTrendingDaysByID(videos):
      La funci n de getMostTrendingDaysByID() itera la lista ordenada y
3
      retorna el elemento que m s se repite.
5
      elemento = lt.firstElement(videos)
6
      mayor_titulo = None
      mayor = 0
9
      i = 0
10
11
      for video in lt.iterator(videos):
          if video['video_id'] == elemento['video_id']:
12
13
              i += 1
14
          else:
              if i > mayor:
15
                  mayor_titulo = elemento
16
                  mayor = i
17
              i = 1
18
              elemento = video
19
20
      if i > mayor:
21
         mayor_titulo = elemento
22
          mayor = i
23
     return (mayor_titulo, mayor)
```

5 Cuarto Requerimiento

El equipo de análisis quiere conocer cuáles son los n videos diferentes con más likes en un país con un tag específico. Recuerde que el campo tags está compuesto por múltiples palabras o frases entre comillas ("") y separadas por el siguiente caracter —, así que el tag requerido puede ser una sub-cadena en el campo.

tag

Y como respuesta debe presentar en consola la siguiente información:

- title
- channel_title
- publish_time
- views
- likes
- dislikes
- tags

Antes de analizar los ordenes de crecimiento del requerimiento, señalaremos las funciones y sus tareas dentro del modelo MVC. Por lo tanto, empezaremos por el view, seguido del controller y finalmente haremos el análisis del orden principalmente en las funciones del model.

Funciones del View

La función del view le pide al usuario ingresar el país y el tag de referencia para posteriormente invocar la función de getVideosByCountryAndTag

Funciones del Controller

Posteriormente, el controller actúa como puente entre el view y el controller invocando su función homónima de getVideosByCountryAndTag

Funciones del Model

La función de getVideosByCountryAndTag usa tres funciones del model:la primera filtra usando los mapas por categoría, la segunda itera para filtrar los vídeos que incluyen un tag específico y la última sortea los vídeos por likes.

```
def getVideosByCountryAndTag(catalog, tag, country):
      La funcion de getVideosByCountryAndTag() filtra los videos por un pa s
3
      y tag espec fico
      sublist = getVideosByCriteriaMap(catalog, 'country', country).get('videos')
6
      sublist2 = getVideosByTag(sublist, tag)
      sorted_list = sortVideos(
9
          sublist2, int(lt.size(sublist2)), 'comparelikes')
  return sorted_list
10
def getVideosByTag(videos, tag):
      La funcion de getVideosByTag() filtra los videos por un tag
3
      espec fico
4
      lista = lt.newList('ARRAY_LIST')
6
      while i <= lt.size(videos):</pre>
9
10
          c_tags = lt.getElement(videos, i).get('tags')
          tagpresence = tag in c_tags
11
12
13
          if tagpresence:
               element = lt.getElement(videos, i)
14
15
              lt.addLast(lista, element)
16
          i += 1
17
18
   return lista
19
def sortVideos(catalog, size, cmp):
      La Funci n sortVideos() organiza los videos de acuerdo
3
      al par metro cmp.
5
      sub_list = lt.subList(catalog, 0, size)
6
      sub_list = sub_list.copy()
      start_time = time.process_time()
8
9
      if cmp == 'cmpVideosByViews':
          sorted_list = ms.sort(sub_list, cmpVideosByViews)
11
12
      if cmp == 'comparetitles':
          sorted_list = ms.sort(sub_list, comparetitles)
13
14
      if cmp == 'comparelikes':
          sorted_list = ms.sort(sub_list, comparelikes)
15
16
      stop_time = time.process_time()
17
18
      {\tt elapsed\_time\_mseg} \; = \; ({\tt stop\_time} \; - \; {\tt start\_time}) * 1000
```

return elapsed_time_mseg, sorted_list

19

6 Funciones Alternas

Con el fin de aprovechar los beneficios de los mapas para disminuir los ordenes de crecimiento, propusimos dos versiones alternas a la función de getMostTrendingDaysByID, en esta versión solo es necesario hacer dos iteraciones y no hace falta sortear la lista lo que disminuiría en teoría el tiempo de ejecución. Sin embargo, teniendo en cuenta el uso de TAD listas, algunas funciones necesitan reiterar la lista o los mapas creados por lo que en la práctica no fue la función más rápida.

```
def getMostTrendingDaysByIDv1(videos):
      ids = mp.newMap(
3
           500,
           maptype='PROBING',
4
          loadfactor=0.8)
      pos = mp.newMap(
          500,
           maptype='PROBING',
           loadfactor = 0.8)
9
10
      ids_list = lt.newList('ARRAY_LIST')
11
      i = 1
13
      while i <= lt.size(videos):</pre>
14
           video_id = lt.getElement(videos, i).get('video_id')
           presence = lt.isPresent(ids_list, video_id) != 0
16
17
18
              n = int(mp.get(ids, video_id).get('value'))
19
               n += 1
20
               mp.put(ids, video_id, n)
21
           else:
22
23
               mp.put(ids, video_id, 1)
               mp.put(pos, video_id, i)
24
25
               lt.addLast(ids_list, video_id)
26
27
28
      mayor = video_id
      repeticiones_mayor = int(mp.get(ids, mayor).get('value'))
29
      mayor = video_id
30
      repeticiones_mayor = int(mp.get(ids, mayor).get('value'))
31
32
33
      for each_id in ids_list:
           repetitions = int(mp.get(ids, each_id).get('value'))
34
35
           if repetitions > repeticiones_mayor:
36
               mayor = each_id
               repeticiones_mayor = repetitions
37
38
      position = mp.get(pos, mayor).get('value')
39
40
      repetitions = mp.get(ids, mayor).get('value')
      result = lt.getElement(videos, position)
41
42
      return (result, repetitions)
43
```

7 Cálculos de Ordenes de Crecimiento

En este apartado haremos los cálculos de los ordenes de crecimiento de cada una de las funciones. Ya que sería negligente analizar las funciones en términos de "N" (siento este el número total de datos) definiremos las siguientes variables y haremos una aproximación en términos de N para su mejor y peor caso:

- P: Es el número total de países cargados por el catálogo que corresponde a 10 países.
- C: Es el número total de categorías cargadas por el catálogo que corresponde a 32 categorías aunque solo 18 categorías son cargadas ya que las otras no corresponden a ningún vídeo.
- NP: Es el tamaño promedio de datos filtrados por país. Su mejor caso es igual a: NP_B y su peor caso es: NP_W
- NC: Es el tamaño promedio de datos filtrados por categoría. Su mejor caso es igual a: NC_B y su peor caso es: NC_W
- NPC: Es el tamaño promedio de una filtración por país y categoría. Su mejor caso es igual a: NPC_B y su peor caso es: NPC_W
- NPT: Es el tamaño promedio de una filtración por país y tags. Su mejor caso es igual a: NPT_B y su peor caso es: NPT_W

Ahora bien, analizaremos paso a paso las funciones que de mayor orden de crecimiento por requerimiento

• Requerimiento 1:

 $\mathbf{getVideosByCountry:}$ Se demora (P), ya que revisa las llaves del catálogo en el apartado de catalog["countries"].

 $\mathbf{getVideosByCategory:}$ Se demora (NP), ya que filtra la lista anterior por categoría para lo cual necesita iterarla.

sortVideos: Al usar merge sort se demora $(NPC \cdot log(NPC))$ en su mejor, peor y caso promedio. **printResultsv1:** Se demora en congruencia con el parámetro "n", es decir (n) porque esas son las veces que itera la lista.

• Requerimiento 2:

 $\mathbf{getVideosByCountry:}$ Se demora (P), ya que revisa las llaves del catálogo en el apartado de catalog["countries"].

sortVideos: Al usar merge sort se demora $(NP \cdot log(NP))$ en su mejor, peor y caso promedio.

 $\mathbf{getMostTrendingDaysByTitle:}$ Se demora (NP) ya que debe iterar toda la lista sorteada para buscar el mayor.

• Requerimiento 3:

 $\mathbf{getVideosByCategory:}$ Se demora (C), ya que revisa las llaves del catálogo en el apartado de catalog["category_id"].

sortVideos: Al usar merge sort se demora $(NC \cdot log(NC))$ en su mejor, peor y caso promedio. getMostTrendingDaysByTitle: Se demora (NC) ya que debe iterar toda la lista sorteada.

• Requerimiento 4:

 $\mathbf{getVideosByCountry:}$ Se demora (P), ya que revisa las llaves del catálogo en el apartado de catalog["countries"]

 $\mathbf{getVideosByTag:}$ Se demora (NPT), ya que filtra la lista anterior por tags para lo cual necesita iterarla.

sortVideos: Al usar merge sort se demora $(NPT \cdot log(NPT))$ en su mejor, peor y caso promedio. **printResultsv2:** Se demora (NPT) en el peor caso ya que debe escoger los mejores videos y asegurarse de no repetirlos iterando la lista de tamaño (NPT).

8 Resultados

En congruencia con lo anterior, haciendo los cálculos pertinentes para cada una de las notaciones en relación a las variables de tamaño definidas anteriormente, obtuvimos los siguientes ordenes de crecimiento para cada uno de los requerimientos:

	Requerimiento 1	Requerimiento 2
Notacion Tilda	$\sim N(15 + P + NP + (NPC \cdot log(NPC) + n)$	$\sim N(10 + P + NP + (NP \cdot log(NP)))$
Big Theta	$\Theta(NP)$	$\Theta(NP \cdot log(NP))$
Big O	$O(NP_W))$	$O(NP_W \cdot log(NP_W))$
Big Omega	$\Omega(NP_B))$	$\Omega(NP_B \cdot log(NP_B))$
	Requerimiento 3	Requerimiento 4
Notacion Tilda	Requerimiento 3 $\sim N(10 + NC + (NC \cdot log(NC))$	Requerimiento 4 $\sim N(20 + P + 2NPT + (NPT \cdot log(NPT))$
Notacion Tilda Big Theta	<u>-</u>	-
	$\sim N(10 + NC + (NC \cdot log(NC)))$	$\sim N(20 + P + 2NPT + (NPT \cdot log(NPT)))$