



Universidad Tecnológica Centroamericana

Facultad de Ingeniería

**CCC308 Concurrencia y Sistemas Distribuidos
(Manual de Usuario de Proyecto)**

Docente: Omar Figueroa

Sección: 458 7:00 a.m.

Presentado por:

Jairo Alejandro Sierra 11811364

Renato David Lizardo Varela 11811148

Fecha:

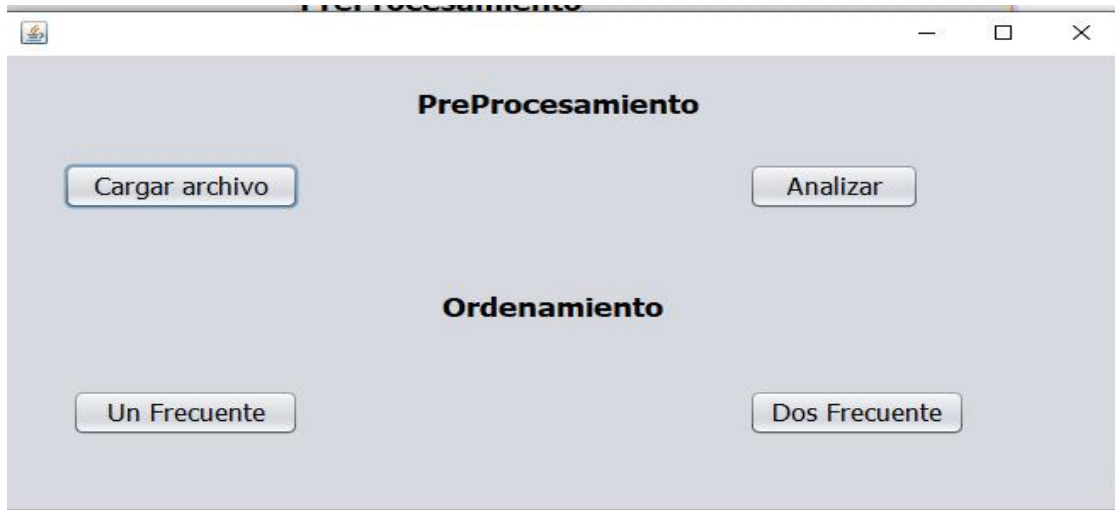
22 de septiembre del 2021

Lugar:

Tegucigalpa, Francisco Morazán, Honduras

PreProcesamiento

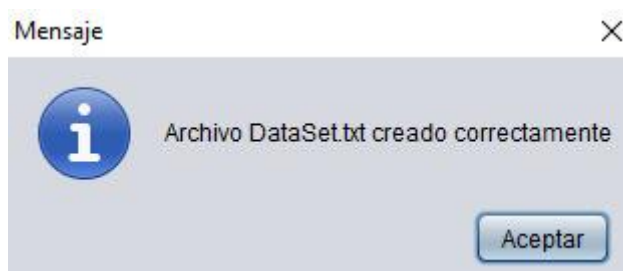
1. Abrir el proyecto de Java llamado PreProcesador, de tal forma ejecutar la clase de Ventana
2. Seguidamente se le mostrara la siguiente ventana:



En donde como primer paso deberá seleccionar el botón de Cargar Archivo, en donde le abrirá una ventana para seleccionar el Dataset original que será preprocesado.



3. Una vez que allá seleccionado se archivo para se preprocesado, deberá seleccionar seguidamente el botón que aparece al lado derecho de la ventana principal llamado Analizar.
4. Una vez que haya terminado el preprocesamiento, se le mostrara la siguiente ventana indicando que el preprocesamiento a finalizado y que el archivo se ha creado correctamente en el directorio principal de la carpeta del proyecto de Java, la ventana seria la siguiente:



Implementación con Hadoop

- Una vez tengamos el documento ya preprocesado, empezaremos a utilizar Hadoop con el MapReduce de palabras Un frecuente y Dos frecuente
- A continuación comando para iniciar los componentes de Hadoop:

```
Administrador: Símbolo del sistema
C:\hadoop\data\namenode\current>fsimage_00000000000000000000, C:\hadoop\data\namenode\current\fsimage_00000000000000000000.md5, C:\hadoop\data\namenode\current\seen_txid, C:\hadoop\data\namenode\current\VERSION]
21/09/17 15:25:19 INFO common.Storage: Storage directory C:\hadoop\data\namenode has been successfully formatted.
21/09/17 15:25:19 INFO namenode.FSImageFormatProtobuf: Saving image file C:\hadoop\data\namenode\current\fsimage.ckpt_00000000000000000000 using no compression
21/09/17 15:25:19 INFO namenode.FSImageFormatProtobuf: Image file C:\hadoop\data\namenode\current\fsimage.ckpt_00000000000000000000 of size 400 bytes saved in 0 seconds .
21/09/17 15:25:19 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
21/09/17 15:25:19 INFO namenode.FSNamesystem: Stopping services started for active state
21/09/17 15:25:19 INFO namenode.FSNamesystem: Stopping services started for standby state
21/09/17 15:25:19 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
21/09/17 15:25:19 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-88T8HT4/192.168.56.1
*****/

C:\WINDOWS\system32>cd C:\hadoop\sbin

C:\hadoop\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\hadoop\sbin>jps
16160 NameNode
8312 Jps
11812 NodeManager
12652 DataNode
13148 ResourceManager

C:\hadoop\sbin>j
```

- Luego de eso, montaremos el dataset preprocesado a la arquitectura de Hadoop con los siguientes comandos:

```
C:\hadoop\sbin>hadoop fs -mkdir /input

C:\hadoop\sbin>hadoop fs -put C:/dataset/DataSet.txt /input

C:\hadoop\sbin>fs -ls /input/
"fs" no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.

C:\hadoop\sbin>hdfs fs -ls /input/
Error: no se ha encontrado o cargado la clase principal fs

C:\hadoop\sbin>hadoop fs -ls /input/
Found 1 items
-rw-r--r--  1 renat supergroup  251516822  2021-09-17  16:39  /input/DataSet.txt

C:\hadoop\sbin>
```

- Una vez montado el Dataset preprocesado en Hadoop, empezaremos a correr el Jar llamado WordCount, el cual nos contara las palabras que salgan con Un Frecuente en la Dataset, el comando sería el siguiente:

```
C:\hadoop\sbin>hadoop jar C:/dataset/WordCount.jar WordCount /input /out
21/09/17 16:45:07 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
21/09/17 16:45:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
21/09/17 16:45:09 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/renat/.staging/job_1631913970226_0001
21/09/17 16:45:09 INFO input.FileInputFormat: Total input files to process : 1
21/09/17 16:45:09 INFO mapreduce.JobSubmitter: number of splits:2
21/09/17 16:45:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1631913970226_0001
21/09/17 16:45:10 INFO mapreduce.JobSubmitter: Executing with tokens: []
21/09/17 16:45:10 INFO conf.Configuration: resource-types.xml not found
21/09/17 16:45:10 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
21/09/17 16:45:11 INFO impl.YarnClientImpl: Submitted application application_1631913970226_0001
21/09/17 16:45:11 INFO mapreduce.Job: The url to track the job: http://DESKTOP-88T8HT4:8088/proxy/application_1631913970226_0001/
21/09/17 16:45:11 INFO mapreduce.Job: Running job: job_1631913970226_0001
```

Arrojando Como resultado que no ha terminado lo siguiente:

```
C:\hadoop\sbin>hadoop jar C:/dataset/WordCount.jar WordCount /input /out
21/09/17 16:45:07 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
21/09/17 16:45:08 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool
ol interface and execute your application with ToolRunner to remedy this.
21/09/17 16:45:09 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/renat/
.staging/job_1631913970226_0001
21/09/17 16:45:09 INFO input.FileInputFormat: Total input files to process : 1
21/09/17 16:45:09 INFO mapreduce.JobSubmitter: number of splits:2
21/09/17 16:45:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1631913970226_0001
21/09/17 16:45:10 INFO mapreduce.JobSubmitter: Executing with tokens: []
21/09/17 16:45:10 INFO conf.Configuration: resource-types.xml not found
21/09/17 16:45:10 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
21/09/17 16:45:11 INFO impl.YarnClientImpl: Submitted application application_1631913970226_0001
21/09/17 16:45:11 INFO mapreduce.Job: The url to track the job: http://DESKTOP-88T8HT4:8088/proxy/application_1631913970
226_0001/
21/09/17 16:45:11 INFO mapreduce.Job: Running job: job_1631913970226_0001
```

Adicional, para hacer el conteo con el Dataset para las palabras con Dos Frecuente, se deberá de correr exactamente el comando anterior pero con Jar llamado DosFrecuente.jar, el comando sería el siguiente:

Hadoop jar C:/dataset/DosFrecuente.jar WordPair /input /OutPares

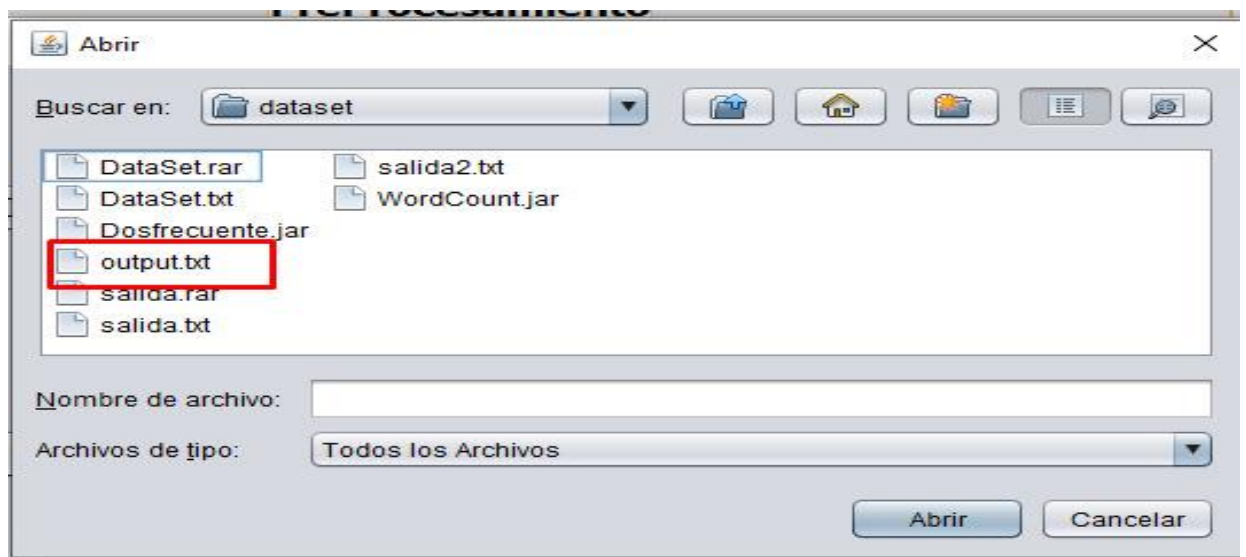
9. Una vez terminado el conteo para cualquiera de los dos Jar, el siguiente comando será para guardarlo en nuestro ordenador un documento con la salida del conteo en formato txt

```
C:\hadoop\sbin>hadoop fs -cat /out/* > C:\dataset\output.txt_
```

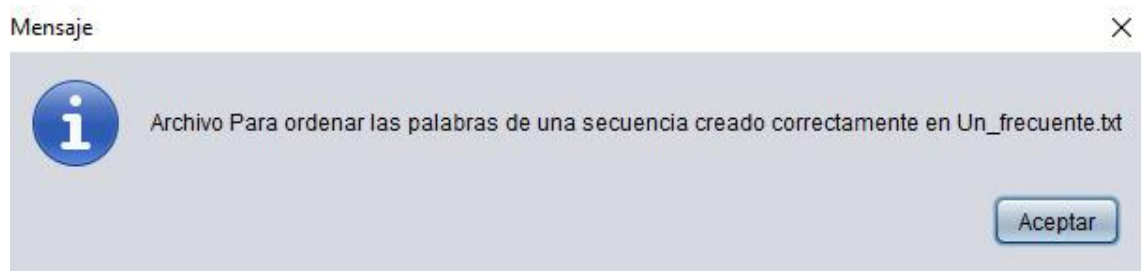
10. Una vez teniendo los dos documentos del WordCount para un frecuente y el WordPair para el de Dos frecuente, comenzaremos a analizar ambos txt y ordenarlos tomando en cuenta las palabras que tengan más de 5000 apariciones
11. El método para analizado del documento de Un Frecuente seria, en la ventana principal del proyecto de Java, se deberá seleccionar el botón llamado Un Frecuente:



Una vez, la aplicación le permitirá elegir el archivo txt que genere en la salida del paso 9 para el conteo de un frecuente con el jar llamado WordCout



Una vez terminado el proceso, la aplicación creará un documento con la salida del análisis de palabras con un frecuente, el cual se guardará en la carpeta principal del proyecto de java, indicando el nombre del txt:



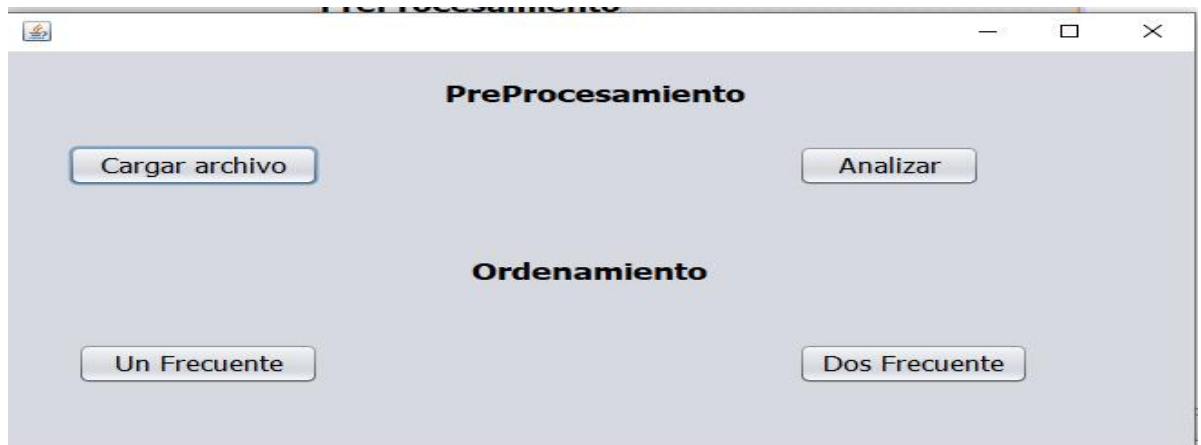
En donde la salida del txt sería la siguiente:

 Un_frecuente.txt: Bloc de notas

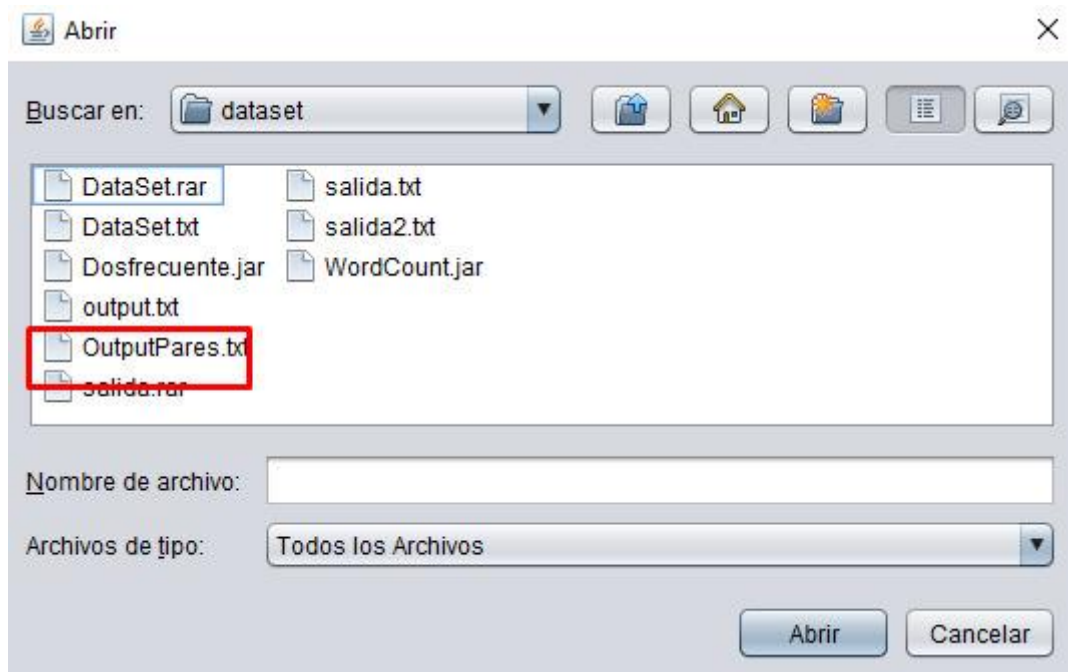
Archivo Edición Formato Ver Ayuda

year 40266
call 40449
last 40940
earnings 42972
price 44072
want 45014
week 48614
long 49496
options 50301
way 50511
today 51363
bought 53575
got 57178
still 58189
see 61098
sell 62176
back 64036
day 65829
stock 66822
make 66881
right 67478
people 73110
time 77874
good 78542
go 79217
one 79795
market 84994
going 88732
puts 89708
now 91863
think 94177
buy 97646
removed 108434
calls 114642
money 115268

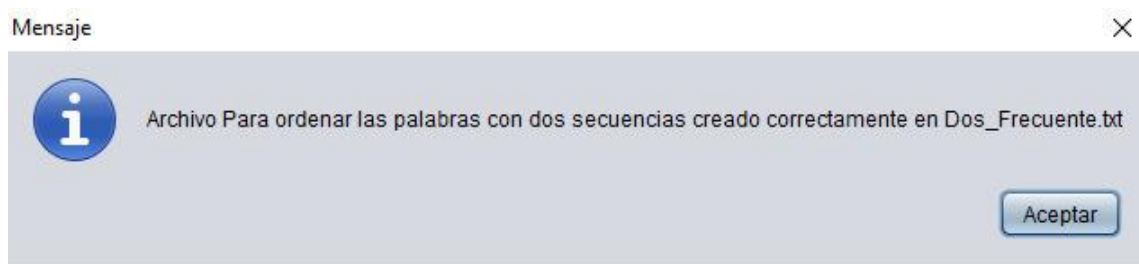
12. El método para analizado del documento de Dos Frecuente seria, en la ventana principal del proyecto de Java, se deberá seleccionar el botón llamado Dos Frecuente:



Una vez, la aplicación le permitirá elegir el archivo txt que se generó en la salida del paso 9 para el conteo de un frecuente con el jar llamado DosFrecuente



Una vez terminado el proceso, la aplicación creara un documento con la salida del análisis de palabras con dos frecuente, el cual se guardará en la carpeta principal del proyecto de java, indicando el nombre del txt:



En donde la salida del txt sería la siguiente:

 Dos_Frecuente.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda

dgaff says 5061
fartbiscuit dgaff 5061
tastes marshmallows 5061
says tastes 5062
concerns moderator 5294
moderators questions 5367
automatically please 5370
contact moderators 5371
questions concerns 5374
please contact 5395
performed automatically 5526
bot action 5530
action performed 5531
stock market 5597
wsbvotebot robot 5708
robot janitor 5713
spaghetti spaghetti 5812
short term 6168
october mwq 6198
somehow lost 6236
puts october 6251
good luck 6612
money puts 6750
buy calls 6794
buy puts 7823
lose money 8105
last week 8158
lost money 8574
1 ec 8710
long term 8927
next week 9523
make money 11176
avgazn retard 12834
right now 21977

Y con esto, tendríamos terminado ambos proceso para el conteo de palabra en nuestro Dataset Original