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:

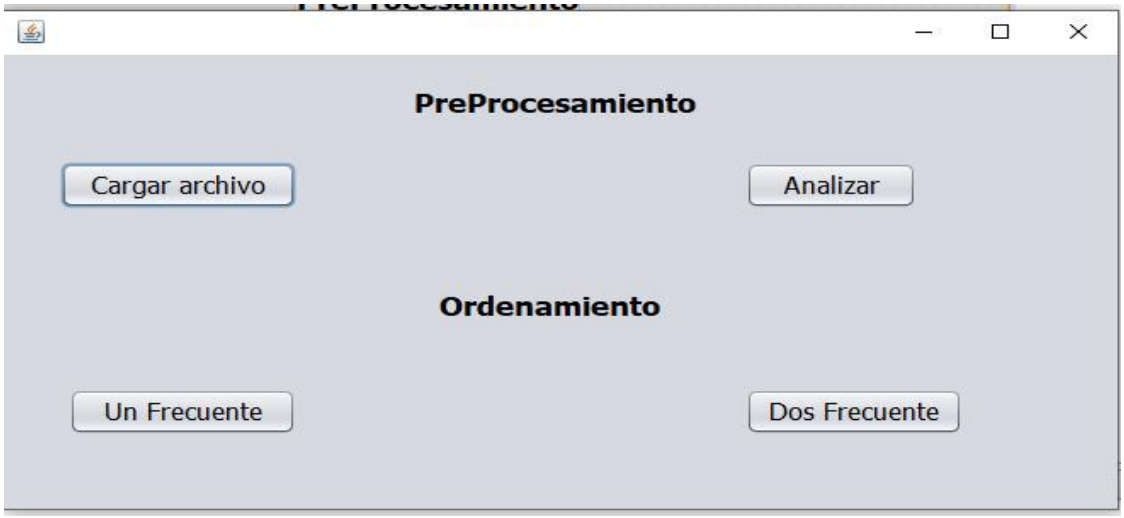
22 de septiembre del 2021

Lugar:

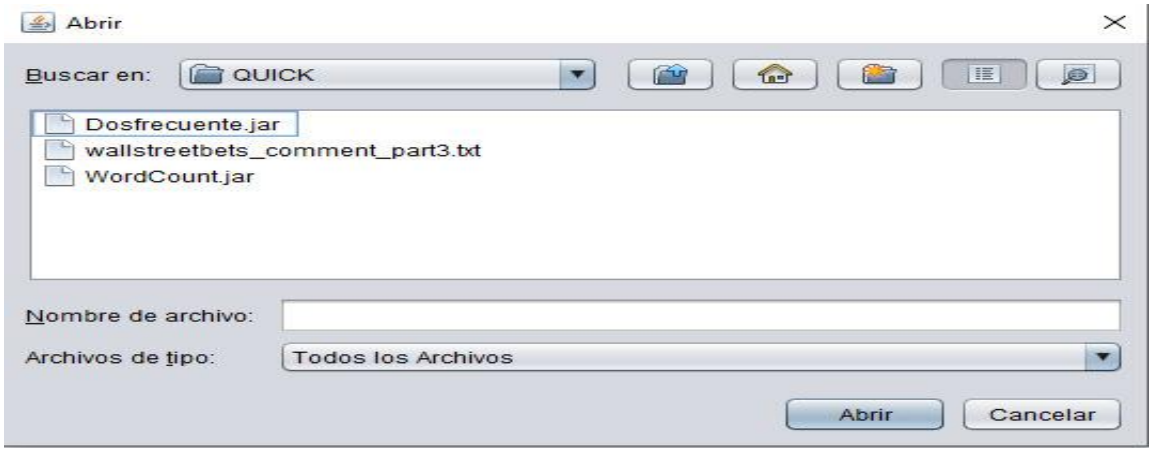
Tegucigalpa, Francisco Morazán, Honduras

PreProcesamiento

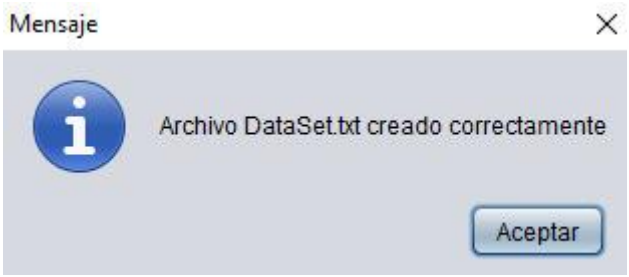
1. Abrir el proyecto de Java llamado PreProcesador, de tal forma ejecutar la clase de Ventana
2. Seguidamente se le mostrara la siguiente ventana:



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.

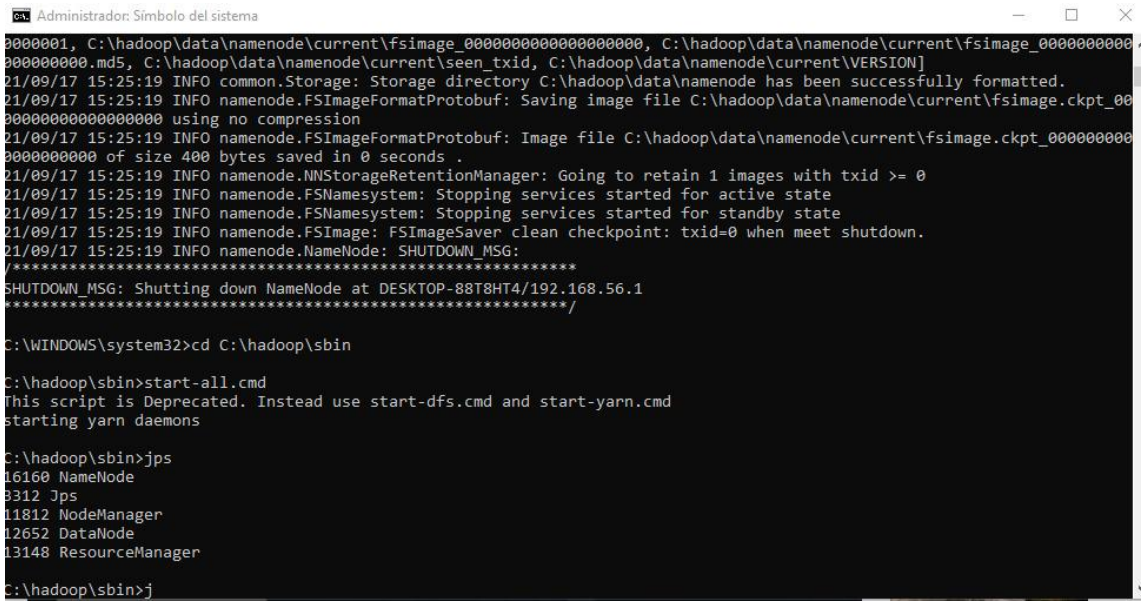


3. Una vez que allá seleccionado se archivo para se preprocesado, deberá seleccionar seguidamente el botón que aparece al lado derecho de la ventana principal llamado Analizar.
4. Una vez que haya terminado el preprocesamiento, se le mostrara la siguiente ventana indicando que el preprocesamiento a finalizado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

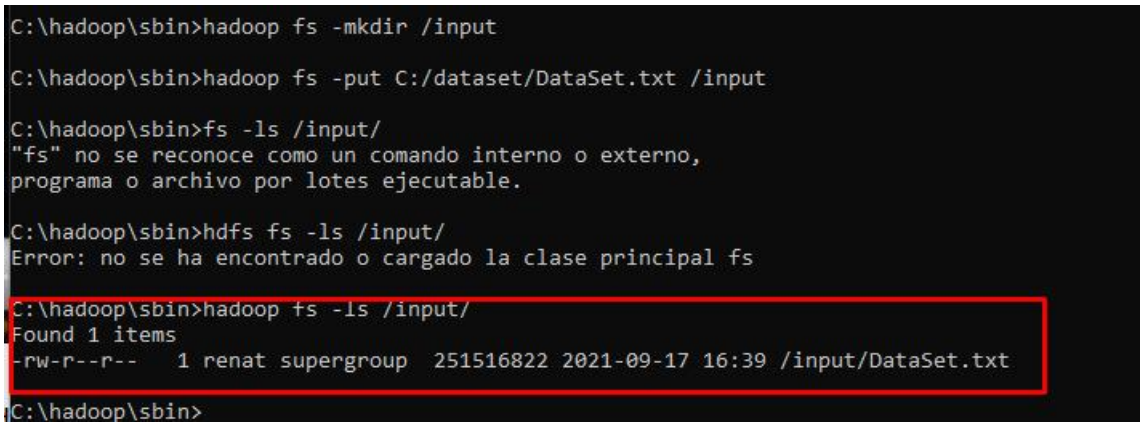


Implementación con Hadoop

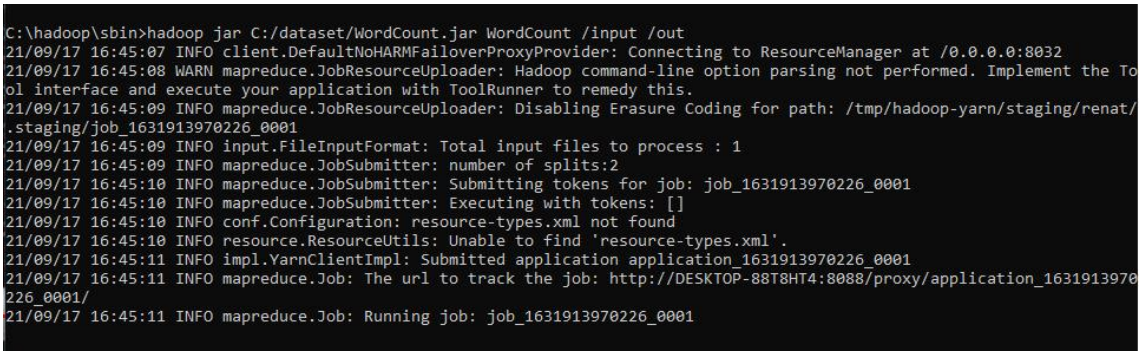
5. Una vez tengamos el documento ya preprocesado, empezáremos a utilizar Hadoop con el MapReduce de palabras Un frecuente y Dos frecuente
6. A continuación comando para iniciar los componentes de Hadoop:



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



8. 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:



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 To
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 DosFrencuente.jar, el comando sería el siguiente:

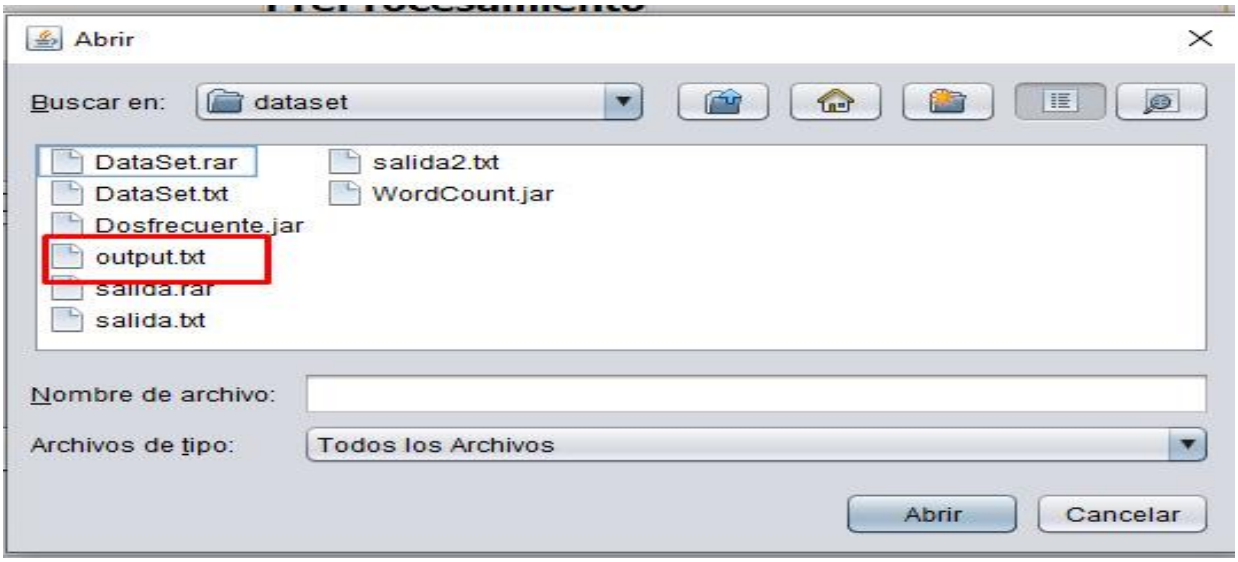
Hadoop jar C:/dataset/DosFrecuente.jar WordPair /input /OutPares

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

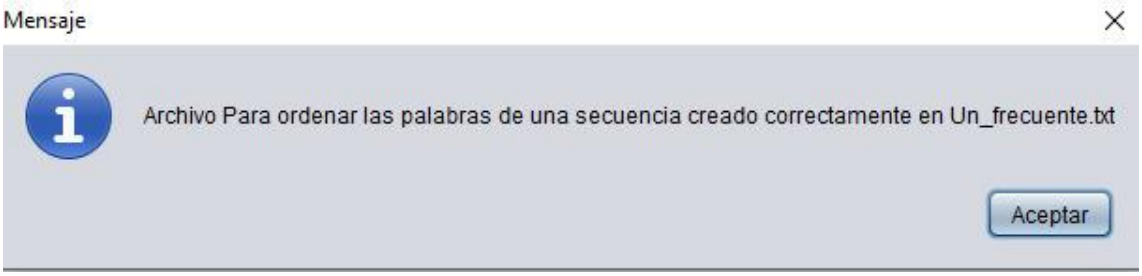
```
C:\hadoop\sbin>hadoop fs -cat /out/* > C:\dataset\output.txt
```

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 genero en la salida del paso 9 para el conteo de un frecuente con el jar llamado WordCout



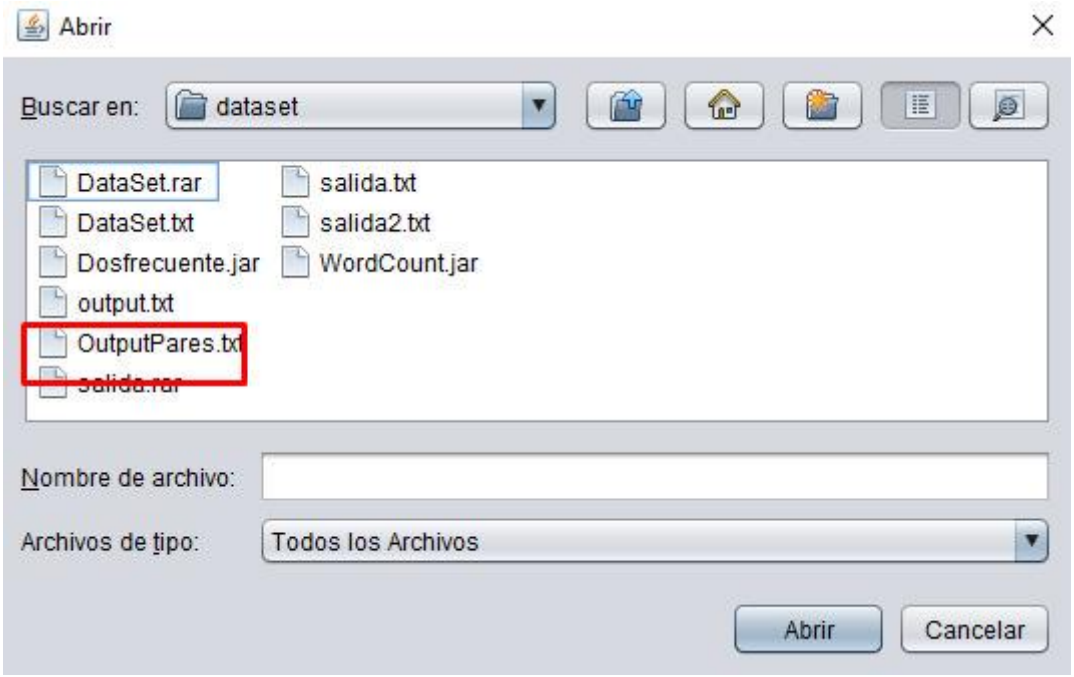
Una vez terminado el proceso, la aplicación creará un documento con la salida del análisis de palabras con un frecuente, el cual se guardará en la carpeta principal del proyecto de java, indicando el nombre del txt:



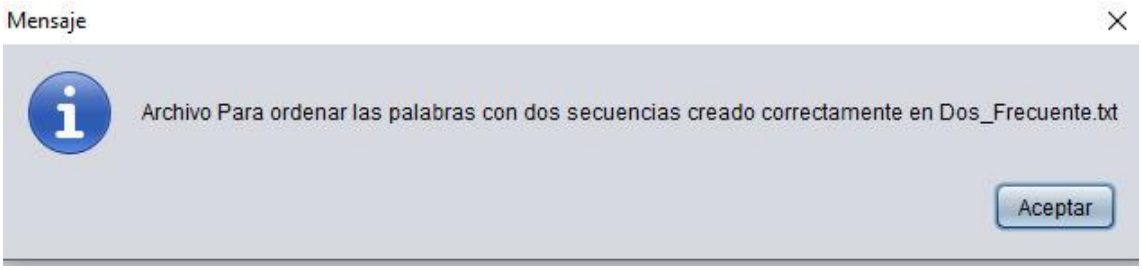
En donde la salida del txt sería la siguiente:

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:

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