**README File**

**PROGRAM 1**

**Statement:** Find how many articles that contain a given specific keyword (either

In the title or either in the content; the keyword is given as the second argument),

Like “texas tech” or “red raider”

**\*\*\* Hadoop Cluster \*\*\***

1. Login to the Hadoop Cluster:

**ssh USERNAME@hadoop.hpcc.ttu.edu**

**ssh hdn1001**

2. Copy the WikiArticleCount Directory which has WikiArticleCount.Java file into your Local Directory using Winscp.

3. Change Directory to WikiArticleCount

**cd WikiArticleCount**

4. Compile the Java code Using Java Compiler:

**Step 1:** **javac -classpath $HADOOP\_CLASSPATH'/\*' Java\_Source\_Code.java**

**Expected Output:**

[yrathams@hdn1001 WikiArticleCount]$ javac -classpath $HADOOP\_CLASSPATH'/\*' WikiArticleCount.java

Note: WikiArticleCount.java uses or overrides a deprecated API.

Note: Recompile with -Xlint:deprecation for details.

**Step 2:**  Create a Jar file.

**Command: jar -cvf Executable\_File\_Name.jar \*.class**

**Expected Output:**

[yrathams@hdn1001 WikiArticleCount]$ jar -cvf Assignment1.jar \*.class

added manifest

adding: WikiArticleCount.class(in = 1930) (out= 1028)(deflated 46%)

adding: WikiArticleCount$IntSumReducer.class(in = 1756) (out= 743)(deflated 57%)

adding: WikiArticleCount$TokenizerMapper.class(in = 1973) (out= 835)(deflated 57%)

5. Execute a Map/Reduce Program:

Argument 1: Input file i.e dataset

Argument 2: Keyword

**Command:** **hadoop jar Assignment1.jar WikiArticleCount /CS5331\_Examples/Programming\_Project\_Dataset.txt "Texas Tech"**

Output:

**Texas Tech 2033**

6. You can also check the Output Using following Command

An output path is displayed at the end of the execution

**Please Check the Path /user/yrathams/WikiArticleCount\_1413412750279 for OutPut**

**Command: hadoop fs -cat /user/yrathams/WikiArticleCount\_1413412750279/\***

**Expected Output:**

[yrathams@hdn1001 WikiArticleCount1]$ hadoop fs –cat /user/yrathams/WikiArticleCount\_1413412750279/\*

**Texas Tech 2033**

**/\*\*\*\*\*\*\*\* Program 2 in Next Page \*\*\*\*\*\*\*\*\*\*\*\*\*\*/**

**PROGRAM 2**

Statement: Twist the Word Count program and develop a new program that finds the five most frequent words in all articles whose title contains the given keyword (the second argument)

**\*\*\* Hadoop Cluster \*\*\***

1. Login to the Hadoop Cluster:

**ssh USERNAME@hadoop.hpcc.ttu.edu**

**ssh hdn1001**

2. Copy the TopFiveWorkdsWiki Directory which has TopFiveWorkdsWiki.java file into your Local Directory using Winscp.

3. Change Directory to TopFiveWorkdsWiki

**cd TopFiveWorkdsWiki**

4. Compile the Java code Using Java Compiler:

**Step 1**: **javac -classpath $HADOOP\_CLASSPATH'/\*' TopFiveWorkdsWiki.java**

**Expected Output:**

[yrathams@hdn1001 TopFiveWorkdsWiki1]$ javac -classpath $HADOOP\_CLASSPATH'/\*' TopFiveWorkdsWiki.java

Note: TopFiveWorkdsWiki.java uses or overrides a deprecated API.

Note: Recompile with -Xlint:deprecation for details.

**Step 2**: Create a jar File

**Command: jar -cvf Executable\_File\_Name.jar \*.class**

[yrathams@hdn1001 TopFiveWorkdsWiki1]$ jar -cvf TopFiveWorkdsWiki.jar \*.class

added manifest

adding: TopFiveWorkdsWiki.class(in = 3873) (out= 1935)(deflated 50%)

adding: TopFiveWorkdsWiki$IntSumCombiner.class(in = 1761) (out= 749)(deflated 57%)

adding: TopFiveWorkdsWiki$IntSumReducer.class(in = 3483) (out= 1462)(deflated 58%)

adding: TopFiveWorkdsWiki$TokenizerMapper.class(in = 2258) (out= 1008)(deflated 55%)

**Step 3**: Provide read – write – execute Permissions to Jar File

Command: chmod 777 **Executable\_File\_Name.jar**

5. Execute a Map/Reduce Program:

Argument 1: Input file i.e dataset

Argument 2: Keyword

**Command: hadoop jar TopFiveWorkdsWiki.jar TopFiveWorkdsWiki /CS5331\_Examples/Programming\_Project\_Dataset.txt "Texas Tech"**

**Expected Output:**

[yrathams@hdn1001 TopFiveWorkdsWiki1]$ hadoop jar TopFiveWorkdsWiki.jar TopFiveWorkdsWiki /CS5331\_Examples/Programming\_Project\_Dataset.txt "Texas Tech"

**Please Check the Path /user/yrathams/TopFiveWorkdsWiki\_1413415259512 for OutPut**

**Output:**

name="cite 923

of 1384

<link 1604

the 2142

Tech 2210

6. Check the Output Using following Command

An output path is displayed at the end of the execution

**Please Check the Path /user/yrathams/TopFiveWorkdsWiki\_1413415259512 for OutPut**

**Command:**

[yrathams@hdn1001 TopFiveWorkdsWiki1]$ hadoop fs -cat /user/yrathams/TopFiveWorkdsWiki\_1413415259512/\*

Output:

name="cite 923

of 1384

<link 1604

the 2142

Tech 2210

**/\*\*\*\*\*\*\*\*\*\*\*\* Program 3 in Next Page \*\*\*\*\*\*\*\*\*\*\*\*\*\*/**

**PROGRAM 3**

Statement: Displays the articles with respect to the search term based on the options like title, date, year and month as argument. Search term being the second argument. This functionality searches and displays the article details based on the arguments. You can select the option through which you want to search the input file.

If the 2nd Argument is “0”: the search term should be a String to search

If the 2nd Argument is “1”: the search term should be in the “YYYY” format

If the 2nd Argument is”2”: the search term should be “YYYY-MM” format

If the 2nd Argument is”3”: the search term should be “YYYY-MM-DD” format

**\*\*\* Hadoop Cluster \*\*\***

1. Login to the Hadoop Cluster:

**ssh USERNAME@hadoop.hpcc.ttu.edu**

**ssh hdn1001**

2. Copy the WikiArticleCount Directory which has WikiArticleSearch.Java file into your Local Directory using Winscp.

3. Change Directory to WikiArticleSearch

**cd WikiArticleSearch**

4. Compile the Java code Using Java Compiler:

Step 1: **javac -classpath $HADOOP\_CLASSPATH'/\*' WikiArticleSearch.Java**

**Expected Output:**

[yrathams@hdn1001 WikiArticleSearch]$ javac -classpath $HADOOP\_CLASSPATH'/\*' WikiArticleSearch.Java

Note: WikiArticleSearch.Java uses or overrides a deprecated API.

Note: Recompile with -Xlint:deprecation for details.

Step 2: Create a Jar file.

**Command: jar -cvf Executable\_File\_Name.jar \*.class**

5. Execute a Map/Reduce Program:

Argument 1: Input file i.e dataset

Argument 2: Option with which you want to search

Argument 3: keyword

**Command:** hadoop jar WikiSearch.jar WikiArticleSearch /CS5331\_Examples/Programming\_Project\_Dataset.txt 0 "Texas Tech"

**Expected Output:**

[yrathams@hdn1001 WikiArticleSearch]$ hadoop jar WikiSearch.jar WikiArticleSearch /CS5331\_Examples/Programming\_Project\_Dataset.txt 0 "Texas Tech"

**Please Check the Path /user/yrathams/WikiArticleSearch\_1413411572990 for OutPut**

Output : Given Below

6. Check the Output Using following Command

An output path is displayed at the end of the execution

**Please Check the Path /user/yrathams/WikiArticleSearch\_1413411572990 for OutPut**

**Expected Output:**

[yrathams@hdn1001 WikiArticleSearch]$ hadoop fs -cat /user/yrathams/WikiArticleSearch\_1413411572990/\*

Texas Tech Texas Tech University

Texas Tech Texas Tech University Goin' Band from Raiderland

Texas Tech Texas Tech University Health Sciences Center

Texas Tech Category:Texas Tech University

Texas Tech List of Texas Tech University people

Texas Tech Category:Texas Tech Red Raiders football

Texas Tech Category:Texas Tech Red Raiders football players

Texas Tech Texas Tech Red Raiders

Texas Tech Category:Wikipedians by alma mater: Texas Tech University

Texas Tech Category:Texas Tech Red Raiders football coaches

Texas Tech Category:Texas Tech basketball

Texas Tech Category:Texas Tech Lady Raiders basketball players

Texas Tech Category:Texas Tech Red Raiders basketball players

Texas Tech Category:Texas Tech Red Raiders basketball

Texas Tech Category:Texas Tech Lady Raiders basketball

Texas Tech Category:Texas Tech Red Raiders basketball coaches

Texas Tech Template:Texas Tech University System

Texas Tech Texas Tech University System

Texas Tech Texas Tech University at Abilene

Texas Tech Texas Tech University at Amarillo

Texas Tech Texas Tech University Center at Junction

Texas Tech Texas Tech University at Fredericksburg

Texas Tech Texas Tech University at Highland Lakes

Texas Tech Texas Tech University Health Sciences Center Amarillo Campus

Texas Tech Texas Tech University Health Sciences Center El Paso Campus

Texas Tech Texas Tech University Health Sciences Center Permian Basin Campus

Texas Tech Category:Texas Tech University System

Texas Tech Category:Texas Tech Red Raiders

Texas Tech Category:Texas Tech University alumni

Texas Tech Category:Texas Tech Red Raiders baseball players

Texas Tech Category:Texas Tech Red Raiders baseball coaches

Texas Tech Category:Texas Tech Red Raiders baseball

Texas Tech Texas Tech University School of Law

Texas Tech Template:Texas Tech Red Raiders football coach navbox

Texas Tech Texas Tech Red Raiders football

Texas Tech Category:Texas Tech University templates

Texas Tech Texas Tech University College of Architecture

Texas Tech Category:Texas Tech University Colleges

Texas Tech Texas Tech University Center in Seville

Texas Tech Texas Tech University Center in Quedlinburg

Texas Tech Texas Technological College Dairy Barn

Texas Tech Texas Tech University Press

Texas Tech Category:Texas Tech University faculty

Texas Tech Category:Texas Tech University people

Texas Tech File:Texas Tech Jones SBC Stadium.JPG

Texas Tech File:Texas Technological College Dairy Barn.JPG

Texas Tech Category:Texas Tech Lady Raiders basketball coaches

Texas Tech Texas Tech University College of Agricultural Sciences & Natural Resources

Texas Tech Texas Tech University College of Mass Communications

Texas Tech Texas Tech University College of Visual & Performing Arts

Texas Tech Texas Tech University Graduate School

Texas Tech Template:Texas Tech Colleges

Texas Tech Texas Tech University College of Education

Texas Tech Texas Tech University College of Human Sciences

Texas Tech Texas Tech University College of Arts & Sciences

Texas Tech Template:User Project Texas Tech University

Texas Tech Template:Texas Tech University

Texas Tech Template:WikiProject Texas Tech University

Texas Tech Category:Wikipedian Texas Tech fans

Texas Tech Texas Tech University Health Sciences Center School of Pharmacy Abilene Campus

Texas Tech 2007 Texas Tech Red Raiders football team

Texas Tech 2006 Texas Tech Red Raiders football team

Texas Tech Texas Tech Red Raiders men's basketball

Texas Tech 2008 Texas Tech Red Raiders football team

Texas Tech Texas Tech Administrative Law Journal

Texas Tech Category:Texas Tech Red Raiders athletic directors

Texas Tech File:SI Texas Tech.jpg

Texas Tech Category:Images of Texas Tech University

Texas Tech Texas Tech Shackleton Glacier Expedition

Texas Tech Template:Texas Tech Red Raiders athletics

Texas Tech File:Texas Tech SET logo.jpg

Texas Tech Texas Tech University Independent School District

Texas Tech Texas Tech Red Raiders baseball

Texas Tech Template:Texas Tech bowl games

Texas Tech Template:Texas Tech baseball coach

Texas Tech Museum of Texas Tech University

Texas Tech Category:Texas Tech Red Raiders football seasons

Texas Tech Texas Tech University traditions

Texas Tech Category:WikiProject Texas Tech University

Texas Tech Category:WikiProject Texas Tech University participants

Texas Tech Template:Invite User Texas Tech University

Texas Tech Category:FA-Class Texas Tech University articles

Texas Tech Category:WikiProject Texas Tech University articles

Texas Tech Category:Texas Tech University articles by quality

Texas Tech Category:Top-importance Texas Tech University articles

Texas Tech Category:High-importance Texas Tech University articles

Texas Tech Category:Mid-importance Texas Tech University articles

Texas Tech Category:Low-importance Texas Tech University articles

Texas Tech Category:Unknown-importance Texas Tech University articles

Texas Tech Category:B-Class Texas Tech University articles

Texas Tech Category:C-Class Texas Tech University articles

Texas Tech Category:Start-Class Texas Tech University articles

Texas Tech Category:Stub-Class Texas Tech University articles

Texas Tech Category:Unassessed Texas Tech University articles

Texas Tech Category:Texas Tech University articles by importance

Texas Tech Portal:Texas Tech University

Texas Tech Portal:Texas Tech University/box-header

Texas Tech Portal:Texas Tech University/box-footer

Texas Tech Portal:Texas Tech University/Intro

Texas Tech Portal:Texas Tech University/Selected article/1

Texas Tech Category:Texas Tech University portal

Texas Tech Portal:Texas Tech University/Projects

Texas Tech Portal:Texas Tech University/Selected picture/4

Texas Tech Portal:Texas Tech University/Did you know

Texas Tech Portal:Texas Tech University/Selected picture/2

Texas Tech Portal:Texas Tech University/Categories

Texas Tech Portal:Texas Tech University/Selected picture/1

Texas Tech Portal:Texas Tech University/Selected picture/3

Texas Tech Portal:Texas Tech University/Wikimedia

Texas Tech Portal:Texas Tech University/Opentask

Texas Tech Portal:Texas Tech University/Related portals

Texas Tech Portal:Texas Tech University/Nominate/Selected picture

Texas Tech Portal:Texas Tech University/News

Texas Tech Portal:Texas Tech University/Selected article/2

Texas Tech Portal:Texas Tech University/Selected article/4

Texas Tech Portal:Texas Tech University/Selected article/5

Texas Tech Portal:Texas Tech University/Selected picture/5

Texas Tech Portal:Texas Tech University/Selected article/3

Texas Tech Portal:Texas Tech University/Nominate/Selected article

Texas Tech Portal:Texas Tech University/Nominate/Selected biography

Texas Tech Portal:Texas Tech University/Selected picture

Texas Tech Portal:Texas Tech University/Selected biography/5

Texas Tech Portal:Texas Tech University/Selected biography/3

Texas Tech Portal:Texas Tech University/Selected biography/2

Texas Tech Portal:Texas Tech University/Selected biography/4

Texas Tech Portal:Texas Tech University/Selected biography/1

Texas Tech Portal:Texas Tech University/Did you know/archive

Texas Tech Portal:Texas Tech University/Selected article

Texas Tech Portal:Texas Tech University/Selected biography

Texas Tech Template:WikiProject Texas Tech University/doc

Texas Tech Category:GA-Class Texas Tech University articles

Texas Tech Category:List-Class Texas Tech University articles

Texas Tech Category:A-Class Texas Tech University articles

Texas Tech Category:NA-importance Texas Tech University articles

Texas Tech Category:Texas Tech University articles

Texas Tech Category:NA-Class Texas Tech University articles

Texas Tech Category:Image-Class Texas Tech University articles

Texas Tech Category:Category-Class Texas Tech University articles

Texas Tech Category:Disambig-Class Texas Tech University articles

Texas Tech Category:Portal-Class Texas Tech University articles

Texas Tech Category:Redirect-Class Texas Tech University articles

Texas Tech Category:Template-Class Texas Tech University articles

Texas Tech Category:Project-Class Texas Tech University articles

Texas Tech Category:Needed-Class Texas Tech University articles

Texas Tech Category:Future-Class Texas Tech University articles

Texas Tech Category:Merge-Class Texas Tech University articles

Texas Tech Portal talk:Texas Tech University

Texas Tech Texas Tech Red Raiders football teams (2000 to 2005)

Texas Tech Category:FL-Class Texas Tech University articles

Texas Tech File:Texas Tech SET logo.png