```
#!/bin/bash
# Fariha's Shell Script for Lab01 Part 3 (Handles Austen, Shakespeare, and Bible)

set -e  # Exit immediately if any command fails

# === SETUP ===
hdfs_base_dir="/lab1"
class_name="WordCountClean"
input_root="/home/ubuntu/textcorpora"
output_root="/home/ubuntu"

# === AUTHORS/DATASETS ===
declare -A datasets
datasets=(
    ["austen"]="austen-*"
    ["shakespeare"]="shakespeare-*"
    ["bible"]="bible-*"
)

# === LOOP OVER EACH CORPUS ===
for author in "${!datasets[@]}"; do
    echo "===== Processing $author ====="

    hdfs_input="$hdfs_base_dir/input-$author"
    hdfs_output="$hdfs_base_dir/output-$author"
    local_merge_output="$output_root/word-count-$author-clean.txt"
    file_pattern="/home/ubuntu/textcorpora/${datasets[$author]}"

    # Create input dir and put files
    hdfs dfs -mkdir -p $hdfs_input
    hdfs dfs -put -f $file_pattern $hdfs_input/

    # Compile
    chmod +x ./compile-map-reduce
    ./compile-map-reduce $class_name

    # Run job
    chmod +x ./run-map-reduce
    ./run-map-reduce $class_name $hdfs_input $hdfs_output

    # Merge output
    hdfs dfs -getmerge $hdfs_output $local_merge_output
    echo "Saved cleaned word count to $local_merge_output"

    # Cleanup HDFS
    hdfs dfs -rm -r -f "$hdfs_input" "$hdfs_output"
done

# === Cleanup Local Files ===
rm -f $class_name.jar *.class
echo "=== DONE ==="
```

```
#!/usr/bin/python3

print("Hello")
def calculate_richness(unique_word_file):
    unique_words = 0
    total_words = 0
    with open(unique_word_file, 'r') as file:
        for line in file:
            parts  = line.strip().split('\t')
            if len(parts) ==2:
                words, count = parts
                unique_words +=1
                total_words += int(count)

    return unique_words/total_words

#Shakespeare
file_shakespeare ='/home/ubuntu/word-count-shakespeare-clean.txt'
richness_ratio = calculate_richness(file_shakespeare)
print(f"Richness of Vocabulary for Shakespeare: {richness_ratio: .4f}")
#Austen
file_austin ='/home/ubuntu/word-count-austen-clean.txt'
richness_ratio = calculate_richness(file_austin)
print(f"Richness of Vocabulary for Austen: {richness_ratio: .4f}")
#King James Bible
file_bible ='/home/ubuntu/word-count-bible-clean.txt'
richness_ratio = calculate_richness(file_bible)
print(f"Richness of Vocabulary for King James bible: {richness_ratio: .4f}")
```

```
ubuntu@ip-172-31-37-140:~$ python3 Vocab.py
Hello
Richness of Vocabulary for Shakespeare:  0.1157
Richness of Vocabulary for Austen:  0.0381
Richness of Vocabulary for King James bible:  0.0202
```