Neustar Coding Task

The below coding assignment is intended to allow us to understand how you approach coding. Since we are looking for someone to write code we should see the code that you write. So here is a coding assignment that we would like you to complete. As part of the assignment, please follow good software engineering practices, demonstrate how you would normally test software, and document and design or coding decisions you have made.

Coding Assignment - Categories/Sets, etc.

Read a file (name passed in as 1st argument to program). Each line of the file has a category, a space, and a sub-category (sub-category can have whitespace within it). Want to:

- Only process the pair (category, sub-category) once
- If a pair appears twice, ignore the second one
- Keep track of the order of the first occurrence of each pair
- Keep track of the count for each category
- Legal category values are: PERSON PLACE ANIMAL COMPUTER OTHER
- Illegal category values should be ignored

Output:

CATEGORY COUNT
PERSON
PLACE
ANIMAL
COMPUTER
OTHER
Ordered list of input
-- Category output should be in the order shown
-- If there are no items in the file for a specified category, it should be shown with a count of 0

Please follow good software engineering practices. In addition to a working program, unit tests should be produced.

Example:

Input file:

```
PERSON Bob Jones
PLACE Washington
PERSON Mary
COMPUTER Mac
```

```
PERSON Bob Jones
OTHER Tree
ANIMAL Dog
PLACE Texas
FOOD Steak
ANIMAL Cat
```

Output:

```
CATEGORY     COUNT
PERSON       2
PLACE        2
ANIMAL       2
COMPUTER     1
OTHER        1
PERSON Bob Jones
PLACE Washington
PERSON Mary
COMPUTER Mac
OTHER Tree
ANIMAL Dog
PLACE Texas
ANIMAL Cat
```

Instructions, Design and Test

Requirements

- Java
- Maven

In console run – mvn install to execute.

Design

A Data class is implemented that stores Category and Sub-Category of each line. Equals is overriden so that objects can be compared and duplicates are avoided when inserting the data objects into a data structure like HashSet. hashCode method of Object is also overridden as it must be overridden always along with equals method. toString is overridden to print the string representation of the object.

A Category enum is implemented to list the valid Categories. The ordinal in the enum will be later used to store the count of the category data in an array. We can store the count in the enum itself using a variable but repeated processing of files in same program will not reset the counts and can cause bugs.

The FileProcessor class implements all the requirements of the application in processFile method and displays the required output using the print functions. LinkedHashSet is used to store the Data objects so that they

can be printed in the order they are read and also to avoid duplicate objects. The count is stored in an integer array with the index value representing the ordinal of the category in Enum. Instead of splitting the string, we extract substrings based on first encountered space since the sub-category can contain white spaces.

Test

Several tests are implemented in Data and FileProcessor classes. Specifically, DataTest class tests the object stores the data correctly, equal implementations is correct and invalid category types are not accepted.

FileProcessorTest creates a test file in run time and run several tests to validate the category counts, invalid categories, duplicate data, invalid file name input and invalid usage of command line arguments.

The tests are not exhaustive but demonstrate the use of JUnit.