2024-01-06

应用系统体系架构 — 作业12

学号：521030990006

姓名：VAHAGN GHAZARYAN

## ***A.Use Spark to implement the same functionality required using Hadoop in the 11th assignment, that is, add the following functionality to your E - BookStore :***

## *i.Store the introductions of all books in your system into multiple text files according to book type.*

## 

## *ii.Write a keyword list that contains several words, for example, [" Java", "JavaScript", "C ++", " Programming", "Star", "Robot "] , etc.*

## 

## *iii.Write a Spark program to read your files into memory and turn them into RDDs, and use the MR job to count the number of times each keyword appears in all book introductions.*

## I use spark’s RDD to run functions. The KeywordCounter class is a Java application that utilizes the Apache Spark framework to perform distributed text processing. It reads text files from a specified directory and counts the occurrences of predefined keywords. The keywords are defined in a list outside the program, and the text files are processed to split their content into words. These words are then filtered against the keyword list, and a tally of each keyword’s occurrence is maintained. *Find the code in spark-app folder.*

## 

## 

## Keep attention on flatMap, filter (**Map**) and countByValue (**Reduce**) functions.

## **RUN**

## First I build the program into jar file and pass it to spark using this command. I also specify keywords I want to count and the file to save the output.

## 

## 

## **Integration with Bookstore**

## I have a controller which listens to some endpoint.

## 

## And a service to run the spark job as command. Note that this is straightforward and simple solution. Later need to figure out better ways like using spark job servers (https://github.com/spark-jobserver/spark-jobserver#development-mode).

## 

## I just have a button and text field to get the input and process the output in my front.

## 

## 

## DEMO

## 