

Programación con R. Práctica 5

Ejercicio 1. Búsqueda y Selección.

Cargar los datos de "student_census.csv" del repositorio haciendo uso de read.csv. Para una celebración deportiva en el campus, se quiere saber si es posible formar un equipo de 5 estudiantes que jueguen al Baloncesto. Para ello se pide localizar a todos los estudiantes que practiquen dicho deporte y de ellos seleccionar 5 mediante una selección aleatoria. Para seleccionarlos se implementará una función que genera números aleatorios uno a uno en el rango adecuado, si el estudiante aún no forma parte del equipo se integrará y si ya está se generará otro número aleatorio hasta que estén seleccionados los cinco.

Ejercicio 2. Ordenación y Regresión.

Con los datos de "student_census.csv" del ejercicio anterior, se pide filtrar en un data.frame las variables Height, Arm.Span, Foot.Size y realizar correlaciones dos a dos. Ordenar los datos de acuerdo con la estatura para determinar la procedencia de los alumnos más altos y hacer un diagrama de barras que indique los resultados por provincias.

Ejercicio 3. MapReduce. Simulación

Con los datos de "student_census.csv" del ejercicio anterior, se pide:

Particionar el dataset en 20 subconjuntos.

Map: Ordenar cada uno de los subconjuntos por estatura en orden decreciente usando el algoritmo quicksort

Implementar una función que mezcle ordenadamente dos subconjuntos previamente ordenados* (también se puede hacer por fuerza bruta, pero no es lo que se pide).

Reduce: Utilizar la función anterior de manera adecuada, el número de veces que se requiera para conseguir la ordenación total del conjunto de 200 registros.

Ejercicio 4. MapReduce. Clasificación

Con los datos de "student_census.csv" del ejercicio anterior, se pide:

Particionar el dataset en 20 subconjuntos.

Map: para cada subconjunto, calcular cuántos estudiantes practican cada uno de los deportes en cada provincia

Reduce: integrar adecuadamente los resultados de cada dataset para determinar cuántos estudiantes de cada provincia practican cada uno de los deportes.

Ejercicio 5. MapReduce. Medias por bloques

Implementar una función que reciba como entrada un dataset de números y un número natural. La función debe devolver la media de los números almacenados en el dataset. Para ello se debe implementar una simulación de la metodología MapReduce de la siguiente forma:

Primero se recogen los datos y el valor n del input

Segundo se divide el dataset en n datasets de tamaños aproximados

Tercero se calcula la suma de cada bloque y su tamaño

Cuarto se combinan los resultados anteriores para obtener la media del data#set original.

Ejercicio 6. Programación Recursiva y Funcional

Observad este código Haskell:

```
> [(l,h)|h<-[0..51], l<-[0..h-1], sum[1..h]==100]
```

Intentar adivinar cuál es el propósito del código y qué resultado obtiene.

Observar esta solución recursiva en R y decidir cuál es el propósito del código:

```
busca <- function (l, h, st){  
  if (l<=st) {if (sum(l:h) ==st) {print(c(l, h))  
                                     return(busca(l+1,l+1,st))}  
    else {if (sum(l:h)<st) return(busca(l, h+1, st))  
          else return(busca(l+1, l+1, st))}  
  }  
}  
  
busca(0,100, 100)
```