

作业4

wordcount

在书上代码基础上进行少量修改，将文本内容全部变为小写，忽略除“-”以外的标点符号

a	36	
able	8	
again	2	
ago	1	
ahead	1	
alabama	3	
all	7	
alleghenies		1
allow	2	
almighty		1
alone	1	
also	1	
always	1	
am	2	
america	5	
american		4
words	3	
work	2	
would	2	
wrongful		1
wrote	1	
years	5	
yes	1	
york	2	
you	8	
your	1	

MatrixMultiply

在源代码基础上少量修改

1, 1	43902
1, 10	38854
1, 11	31510
1, 12	38810
1, 13	41441
1, 14	28892
1, 15	33590
1, 16	28245
1, 17	36005
1, 18	24921
1, 19	43586
1, 2	33480
1, 20	44208
1, 3	43804
1, 4	47429
1, 5	33662
1, 6	27773
1, 7	39023
1, 8	43199
1, 9	22069
10, 1	52471
10, 10	45764

9, 19	38380
9, 2	46772
9, 20	49907
9, 3	44007
9, 4	51999
9, 5	37381
9, 6	28497
9, 7	42048
9, 8	38520
9, 9	30185

关系代数

- Selection

Map阶段，按行判断是否符合条件，若满足则输出得键值对为 (record, NullWritable.get())

无需Reduce阶段

1. 条件为age=18时输出结果

```
1,tom,18,60
4,tony,18,62
7,brown,18,65
12,ivy,18,58
13,sam,18,67
16,steven,18,60
19,coco,18,55
```

2. 条件为age>18时输出结果

```
5,bob,20,65
6,leon,19,58
9,cici,19,55
14,owen,20,63
15,bill,19,62
```

• Projection

map获取属性值

reduce合并输出(name, NullWritable.get())

对名字投影

```
bill  
bob  
brown  
candy  
cici  
coco  
grace  
henry  
ivy  
jack  
jimmy  
leon  
lily  
lucas  
owen  
sam  
steven  
tom  
tony  
zoey
```

- **Union**

map阶段读取两个文本，记录相同的

reduce阶段将所有key输出

求并集

1,tom,18,60
10,grace,16,56
11,henry,17,61
12,ivy,18,58
13,sam,18,67
14,owen,20,63
15,bill,19,62
16,steven,18,60
17,jimmy,16,62
18,lucas,17,59
19,coco,18,55
2,jack,16,59
20,zoey,17,56
21,linda,19,60
22,tina,17,56
25,monica,20,61
3,lily,17,58
4,tony,18,62
5,bob,20,65
6,leon,19,58
7,brown,18,65
8,candy,17,56
9,cici,19,55

- Intersection

求交集

10,grace,16,56
12,ivy,18,58
19,coco,18,55
20,zoey,17,56
3,lily,17,58
8,candy,17,56
9,cici,19,55

- Difference

求差集

```
1,tom,18,60
11,henry,17,61
13,sam,18,67
14,owen,20,63
15,bill,19,62
16,steven,18,60
17,jimmy,16,62
18,lucas,17,59
2,jack,16,59
4,tony,18,62
5,bob,20,65
6,leon,19,58
7,brown,18,65
```

- **NatureJoin**

map阶段输出<id, others>, value为除了id的属性值

reduce合并输出

Ra和Rb在属性id上进行自然连接

```
1,tom,18,60,1,178
2,jack,16,59,1,175
3,lily,17,58,0,165
4,tony,18,62,1,173
5,bob,20,65,1,179
6,leon,19,58,1,180
7,brown,18,65,1,182
8,candy,17,56,0,166
9,cici,19,55,0,168
10,grace,16,56,0,170
11,henry,17,61,1,173
12,ivy,18,58,0,162
13,sam,18,67,1,182
14,owen,20,63,1,177
15,bill,19,62,1,177
16,steven,18,60,1,175
17,jimmy,16,62,1,178
18,lucas,17,59,1,183
19,coco,18,55,0,160
20,zoey,17,56,0,168
```

ps完整代码和运行结果上传github<https://github.com/BublyDong/homework4>