




spark实验感想

原创

[想守护令人心疼的美好](#)  于 2020-11-19 23:30:18 发布  1005  收藏

文章标签: [spark](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/qq_43661545/article/details/109829879

版权

实验三尽力了, map和flatMap的区别还是需要注重以下, 有时候map之后的数据类型不行但是flatMap扁平化就可以, 还有待提高。

另外学会了ctrl+q在intelli可以查看变量的数据类型

实验5和实验6还是有些问题, 还是再更改学习以下吧。

```
val wordsRdd2 = rdd3.flatMap(x => {  
    val wordtext2 = new JiebaSegmenter().sentenceProcess(x._2+x.3).toArray()  
    .map(x => (x.toString, 1)).groupBy(_._1).map(t=>{  
        (t._1,t.2.size)).toList.sortBy(_._2)  
    })  
    wordtext2  
})
```

分词操作附1, 计算单词在每个文章中出现的次数