

实验二:倒排索引

171860607

白晋斌

810594956@qq.com

组号:2020st32

小组成员:

171860607 白晋斌

171860033 黄诗涵

171180541 祖东珏

分工:因本次实验较为简单,小组成员决定分别独立完成全部必做与两部分选做,最后由白晋斌完成实验报告及集群测试,报告中所有截图与代码均以白晋斌的测试为例. 此外,三位同学的源代码、JAR 包、JAR 包执行方式说明详见附件.

目录

实验内容与要求	3
1.实验任务	3
2.输出格式	3
3.选做内容	3
4.提交要求	4
实验过程	4
0.配置环境	4
1.Map 和 Reduce 的设计思路.....	4
2.Map 和 Reduce 的伪代码	4
3.本地测试	5
4.选做内容 1	5
4.1 设计思路	5
4.2 本地测试	5
5.选做内容 2	6
5.1 设计思路	6
5.2 本地测试	6
集群运行	6
1.Jar 包执行方式说明.....	6
2.执行报告内容.....	6
3.输出结果文件.....	8
4.“江湖”“风雪”输出结果	10
4.1 基础任务	10
4.2 拓展任务 1	12
4.3 拓展任务 2	12
参考资料	12

实验内容与要求

1.实验任务

请实现课堂上介绍的“带词频属性的文档倒排算法”。

在统计词语的倒排索引时，除了要输出带词频属性的倒排索引，还请计算每个词语的“平均出现次数”（定义见下）并输出。请在实验报告中注明小组分工的情况。

“平均出现次数”在这里定义为：

$$\text{平均出现次数} = \frac{\text{词语在全部文档中出现的频数}}{\text{包含该词语的文档数}}$$

假如文档集中有四本小说：A、B、C、D。词语“江湖”在文档 A 中出现了 100 次，在文档 B 中出现了 200 次，在文档 C 中出现了 300 次，在文档 D 中没有出现。则词语“江湖”在该文档集中的“平均出现次数”为 $(100 + 200 + 300) / 3 = 200$ 。

注意 这两个计算任务请在同一个 MapReduce Job 中完成。

2.输出格式

对于每个词语，输出一个键值对，该键值对的格式如下：

[词语] \TAB 平均出现次数, 小说 1:词频; 小说 2:词频; 小说 3:词频; ...; 小说 N:词频

输出中的小说名需要去掉“.txt.segmented”的文件名后缀。

下图展示了输出文件的一个片段（图中内容仅为格式示例）：

```
江湖 98.98, 金庸02雪山飞狐:43; 金庸04天龙八部:55; 金庸07鹿鼎记:123; ...  
解药 42, 金庸12倚天屠龙记:41; 金庸15越女剑:45; ...
```

3.选做内容

1.使用另外一个 MapReduce Job 对每个词语的平均出现次数进行全局排序，输出排序后的结果。

2.为每位作家、计算每个词语的 TF-IDF。TF 定义为某个词语在某个作家的所有作品中的出现次数之和。IDF 定义为：

$$\text{IDF}(\text{词语}) = \log\left(\frac{\text{语料库文档总数}}{\text{包含该词的文档数} + 1}\right)$$

输出格式：作家名字，词语，该词语的 TF-IDF。

4.提交要求

在最后提交的压缩包中，除了包含源代码、JAR 包、JAR 包执行方式说明，还需要包含一个实验报告。实验报告中请包含：

1. Map 和 Reduce 的设计思路（含 Key、Value 类型）。
 2. MapReduce 中 Map 和 Reduce 的伪代码（或者带注释的实际代码，如果使用实际代码，请做好排版）。
 3. 输出结果文件的部分截图。输出结果文件在 HDFS 上的路径（某些情况下助教会检查 HDFS 上的输出文件）。
 4. “江湖”、“风雪”两个单词的输出结果。
 5. 请在报告中包含在集群上执行作业后，Yarn Resource Manager 的 WebUI 执行报告内容。请完整包括执行报告内容，否则影响分数。每个 MapReduce Job 对应一个报告）。执行报告内容示例见下文。
- 额外注意的是，本次实验均保留了两位小数。

实验过程

0.配置环境

首先要在 Idea 导入对应的 Jar 包[1]. 主要包括：

hadoop-2.7.1/share/hadoop/mapreduce 下的所有 jar 包

hadoop-2.7.1/share/hadoop/common 下的 hadoop-common-2.7.1.jar

hadoop-2.7.1/share/hadoop/common/lib 下的所有 jar 包

1.Map 和 Reduce 的设计思路

Map:对于输入的文件每句进行切割,将单词,文件名作为 key,设置词频为 1(value),在每次 Mapper 结束后对当前得到的数据进行合并,即调用 Combiner 函数,将 key 相同的 value 进行累加,得到新的 key, value 对.

Partition 用个小技巧将 key 中的 word 分出来作为划分标准,便于将同一单词划分到同一个节点.

Reduce: 将 key 相同的 value 进行累加,得到单词,文件名:词频对,对每个 key 进行分割,得到单词并累积统计其对应的 value,最后除以总文件数得到平均出现次数,最后以单词作为 key,将词频和得到单词,文件名:词频对拼接作为 value 输出.

2.Map 和 Reduce 的伪代码

伪代码如下,详细代码见 `BigDataLab2.java`.

```
class Mapper
procedureMap(docid dn, doc d)
    F ← new AssociativeArray
    for all term t ∈ doc d do
        F{t} ← F{t} + 1
    for all term t ∈ F do
        Emit(term t, posting <dn, F{t}>)
class Reducer
procedureReduce(term [<t, dn>...], postings f...)
    P ← new List
    sumFile = |dn| // |A| means num of A
```

```

sumWord=0
for all posting f ∈ postings do
    Append(P, <dn, f>)
    sumWord+=f
avg=sumWord/sumFile
postings V=strcat(t,avg,P)
Emit(term t, postings V)

```

3.本地测试

构造生成 jar 包[2][3], 导入 hdfs 系统.

```
hadoop fs -rm -r /lab2
```

```
zip -d JavaProJ1.jar META-INF/LICENSE;
```

```
jar tvf JavaProJ1.jar | grep -i license;
```

```
zip -d JavaProJ1.jar LICENSE [4]
```

```
hadoop fs -copyFromLocal ~/Desktop/JavaProJ1.jar /
```

```
hadoop fs -copyFromLocal ~/Desktop/novels/ /
```

```
hadoop jar JavaProJ1.jar BigDataLab2.BigDataLab2 /novels /output
```

```
hadoop fs -copyToLocal /output ~/Desktop/
```

4.选做内容 1

4.1 设计思路

将倒排索引的输出结果作为输入, 在 map 阶段, 将其中的 value 中的平均值作为新的 key, 将原来的 key 和 value 拼接在一起作为新的 value; 在 reduce 阶段, 进行还原, 将 value 中的 key 和 value 取出来作为新的 key 和 value.

值得注意的是, 这里要求的是全排序, 全排序主要有 3 种实现方法. (1). 只定义一个 reduce, 默认就是全排序. (2). 自定义分区函数(自行设置分界区间). (3). 使用 hadoop 的采样机制(即调用 Hadoop 自带的 TotalOrderPartitioner). 这里我们采用第三种实现方式, 先采样得到 partitionFile, 根据 partitionFile 进行 TotalOrderPartitioner.

伪代码如下:

```

class Mapper
    procedureMap(key word, value [avg,<filename1,count1>,...])
        Emit(avg, [word, <filename1,count1>,...])
class Reducer
    procedureReduce(key avg, value [word, <filename1,count1>,...])
        Emit(word, [avg,<filename1,count1>,...])

```

4.2 本地测试

修改主类[7], 或在程序运行前输入:

```
hadoop jar JavaProJ1.jar BigDataLab2.BigDataLab2_1 /output/part-r-00000
/output2 /partitionFile
```

这里遇到了一个错误 org.apache.hadoop.io.LongWritable cannot be cast to org.apache.hadoop.io.Text, 原因见[6]. 解决方法: 设置 job 输入格式.

```
job.setInputFormatClass(KeyValueTextInputFormat.class);
```

此外, 我们原本打算自己编写 mapreduce 函数进行全局排序, 但发现 MapReduce 程序在处理数据的过程中会对数据排序(map 输出的 kv 对传输到 reduce 之前会排序), 排序的依据是 map 输出的 key. 所以我们如果要想实现自己需要的排序规则, 则可

以考虑将排序因素放到 key 中, 让 key 实现接口:WritableComparable, 然后重写 key 的 compareTo 方法来指定比较规则, 这里我们直接使用默认对 key 升序排序的方式. [8]

终端输入: `hadoop fs -copyToLocal /output2 ~/Desktop/`, 从而查看实验结果.

5.选做内容 2

5.1 设计思路

通过 `FileSystem.listStatus` 获取总文档数[9], 之后通过 `Configuration.set()` 在各个 Mapper 之间传递. 将倒排索引的输出结果作为输入, 在 map 阶段, 提取 value 中所有的作者名同时包含该词语的文件进行计数, 作者名作为字典的关键词, 而对应书籍的词频之和作为值, 得到每个作者每个单词的词频后计算出 TF-IDF. 最后, 将作者名作为 Key, 词语#TF-IDF 作为 value 输出.

伪代码如下.

```
class Mapper
  procedureMap(key word, value [avg,<filename1,count1>,...])
    String fileNumber = context.getConfiguration().get("fileNumbers");
    author,bookname=filename.split()
    IDF=log(|filename|/(|value|+1)) //|A| means num of A
    TF=0
    for all [word,author] ∈ [key,value] do
      TF{[word,author]}+=count
    for all term t ∈ F do
      Emit(author, [word,TF*IDF])
```

5.2 本地测试

`hadoop jar JavaProj1.jar BigDataLab2.BigDataLab2_2 /output5/part-r-00000 /output12 /novelsWITHsegmented`

值得注意的是, 第一个参数为输入, 第二个参数为输出, 第三个参数为初始文件地址(方便统计文件数目).

集群运行

1.Jar 包执行方式说明

基本任务:

`hadoop jar /home/2020st32/JavaProj1.jar BigDataLab2.BigDataLab2 /data/exercise_2 /user/2020st32/output_0`

选做任务 1:

`hadoop jar /home/2020st32/JavaProj1.jar BigDataLab2.BigDataLab2_1 /user/2020st32/output_0 /user/2020st32/output_1 /user/2020st32/partitionFile`

选做任务 2:

`hadoop jar /home/2020st32/JavaProj1.jar BigDataLab2.BigDataLab2_2 /user/2020st32/output_0 /user/2020st32/output_2 /data/exercise_2`

2.执行报告内容

集群监控页面:

Show 20 entries											Search: st32	
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes	
application_1572597966684_3151	2020st32	optional-2	MAPREDUCE	root.team32	Tue Apr 21 13:53:31 +0800 2020	Tue Apr 21 13:54:02 +0800 2020	FINISHED	SUCCEEDED	<div></div>	History	N/A	
application_1572597966684_3150	2020st32	optional-1	MAPREDUCE	root.team32	Tue Apr 21 13:51:41 +0800 2020	Tue Apr 21 13:52:12 +0800 2020	FINISHED	SUCCEEDED	<div></div>	History	N/A	
application_1572597966684_3149	2020st32	inverted index	MAPREDUCE	root.team32	Tue Apr 21 13:46:35 +0800 2020	Tue Apr 21 13:48:58 +0800 2020	FINISHED	SUCCEEDED	<div></div>	History	N/A	

Job 执行情况页面



Logged in as: dr.who

Application application_1572597966684_3149

Cluster

[About](#)
[Nodes](#)
[Node Labels](#)
[Applications](#)
[NEW](#)
[NEW SAVING](#)
[SUBMITTED](#)
[ACCEPTED](#)
[RUNNING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)
[Scheduler](#)

Tools

Kill Application

Application Overview

User: 2020st32
Name: inverted index
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: FINISHED
Queue: root.team32
FinalStatus Reported by AM: SUCCEEDED
Started: Tue Apr 21 13:46:35 +0800 2020
Elapsed: 2mins, 23sec
Tracking URL: History
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 1350481 MB-seconds, 1039 vcore-seconds

Show 20 entries					Search:
Attempt ID	Started	Node	Logs	Blacklisted Nodes	
appattempt_1572597966684_3149_000001	Tue Apr 21 13:46:35 +0800 2020	http://slave005:8042	Logs	N/A	

Showing 1 to 1 of 1 entries

FirstPrevious1NextLast



Logged in as: dr.who

Application application_1572597966684_3150

Cluster

[About](#)
[Nodes](#)
[Node Labels](#)
[Applications](#)
[NEW](#)
[NEW SAVING](#)
[SUBMITTED](#)
[ACCEPTED](#)
[RUNNING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)
[Scheduler](#)

Tools

Kill Application

Application Overview

User: 2020st32
Name: optional-1
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: FINISHED
Queue: root.team32
FinalStatus Reported by AM: SUCCEEDED
Started: Tue Apr 21 13:51:41 +0800 2020
Elapsed: 31sec
Tracking URL: History
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 101440 MB-seconds, 59 vcore-seconds

Show 20 entries					Search:
Attempt ID	Started	Node	Logs	Blacklisted Nodes	
appattempt_1572597966684_3150_000001	Tue Apr 21 13:51:41 +0800 2020	http://slave012:8042	Logs	N/A	

Showing 1 to 1 of 1 entries

FirstPrevious1NextLast



Logged in as: dr.who

Application application_1572597966684_3151

Cluster

About
Nodes
Node Labels
Applications
NEW
NEW SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler

Tools

Kill Application

User: 2020st32
Name: optional-2
Application Type: MAPREDUCE
Application Tags:
YarnApplicationState: FINISHED
Queue: root.team32
FinalStatus Reported by AM: SUCCEEDED
Started: Tue Apr 21 13:53:31 +0800 2020
Elapsed: 30sec
Tracking URL: History
Diagnostics:

Application Overview

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 99341 MB-seconds, 58 vcore-seconds

Show 20 entries

Attempt ID

Started

Node

Logs

Blacklisted Nodes

appattempt_1572597966684_3151_000001

Tue Apr 21 13:53:31 +0800 2020

http://slave005:8042

Logs

N/A

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

以下是通过 19888 查看 JobHistory 的截图

2020.04.21 13:46:35 CST	2020.04.21 13:46:41 CST	2020.04.21 13:48:59 CST	job_1572597966684_3149	inverted index	2020st32	root.team32	SUCCEEDED	218	218	1	1
2020.04.21 13:51:41 CST	2020.04.21 13:51:45 CST	2020.04.21 13:52:11 CST	job_1572597966684_3150	optional-1	2020st32	root.team32	SUCCEEDED	1	1	1	1
2020.04.21 13:53:31 CST	2020.04.21 13:53:37 CST	2020.04.21 13:54:02 CST	job_1572597966684_3151	optional-2	2020st32	root.team32	SUCCEEDED	1	1	1	1

3.输出结果文件

如下图所示, output_0, output_1, output_2 分别对应必做任务, 选做 1, 选做 2.

HDFS

Upload New

Actions Move to trash

Home /user/2020st32

Trash

<input type="checkbox"/>	Name	Size	User	Group	Permissions	LastModified
<input type="checkbox"/>	.Trash	0B	2020st32	supergroup	drwxr-xr-x	Tue Apr 21 2020 12:58:18 GMT+0800 (China Standard Time)
<input type="checkbox"/>	output_0	0B	2020st32	supergroup	drwxr-xr-x	Tue Apr 21 2020 13:48:58 GMT+0800 (China Standard Time)
<input type="checkbox"/>	output_1	0B	2020st32	supergroup	drwxr-xr-x	Tue Apr 21 2020 13:52:12 GMT+0800 (China Standard Time)
<input type="checkbox"/>	output_2	0B	2020st32	supergroup	drwxr-xr-x	Tue Apr 21 2020 13:54:02 GMT+0800 (China Standard Time)
<input type="checkbox"/>	partitionFile	138B	2020st32	supergroup	-rw-r--r--	Tue Apr 21 2020 13:51:39 GMT+0800 (China Standard Time)

Show 10 of 5 items

Page 1 of 1

0 1.69,卧龙生07.飞燕惊龙:1;卧龙生37.天涯情侣:1;卧龙生42.新仙鹤神针:1;卧龙生45.燕子传奇:1;李凉07.赌棍小狂侠:5;李凉15.江湖一担皮:1;李凉21.六宝江湖行:1;李凉23.妙贼丁小勾:3;李凉27.奇神杨小邪:1;李凉38.笑笑江湖:4;梁羽生01.白发魔女传:1;梁羽生11.广陵剑:1;梁羽生12.瀚海雄风:1;梁羽生25.牧野流星:3;梁羽生34.武当一剑:1;金庸07.鹿鼎记:1

007 1.00,李凉12.活宝小淘气:1

01 1.00,卧龙生45.燕子传奇:1;李凉23.妙贼丁小勾:1

01章 1.00,古龙60.神君别传:1

02章 1.00,古龙60.神君别传:1

03章 1.00,古龙60.神君别传:1

04 1.00,李凉26.奇神杨小邪续集:1

04章 1.00,古龙60.神君别传:1

05 1.00,李凉23.妙贼丁小勾:1

05章 1.00,古龙60.神君别传:1

06 1.00,李凉23.妙贼丁小勾:1

06章 1.00,古龙60.神君别传:1

07章 1.00,古龙60.神君别传:1

Cancel

Download

慈溪县 1.00,李凉40.新蜀山剑侠传续:1

切送 1.00,梁羽生32.随笔集: 笔花六照:1

慈母手中线 1.00,卧龙生04.地狱门:1;古龙24.护花铃:1;金庸09.书剑恩仇录:1

切错 1.00,李凉04.本尊分身:1;李凉24.妙贼丁小勾续集:1

切除 1.00,古龙04.边城刀声:1;李凉26.奇神杨小邪续集:1;梁羽生17.江湖三女侠:1

怯生 1.00,卧龙生32.天鹤谱:1;古龙24.护花铃:1;古龙29.剑毒梅香:1;古龙31.剑气书香:1;李凉19.酒狂任小赌:1;梁羽生22.龙凤宝钗缘:1

切音 1.00,金庸04.天龙八部:1;金庸12.倚天屠龙记:1

切题 1.00,古龙58.情人箭:1;梁羽生19.狂侠天娇魔女:1;梁羽生30.散花女侠:1

刈 1.00,卧龙生48.银月飞霜:1

慈孝 1.00,卧龙生30.素手劫:1

前呼后应 1.00,卧龙生41.无形剑:1;李凉37.小鱼吃大鱼:1;李凉41.杨小邪发威:1

刊了 1.00,梁羽生32.随笔集: 笔花六照:1

刊到 1.00,李凉25.魔手邪怪:1

桂陈 1.00,卧龙生21.妙绝天香:1;卧龙生26.情剑无刃:1

Cancel

Download

卧龙生 水火之中 8.57

卧龙生 传出 67.38

卧龙生 水火土 17.96

卧龙生 水火无情 343.86

卧龙生 水火棍 4.29

卧龙生 胯下之辱 13.06

卧龙生 传出去 55.74

卧龙生 水灵灵 10.32

卧龙生 水灾 5.10

卧龙生 水烟 8.96

卧龙生 胯下 29.70

卧龙生 鹅蛋脸 3.19

卧龙生 水烟筒 4.69

卧龙生 水烟袋 41.18

卧龙生 水者 5.22

Cancel

Download

4.“江湖”“风雪”输出结果

4.1 基础任务

4.1.1 江湖

江湖 116. 06, 卧龙生 01. 镖旗:275;卧龙生 02. 春秋笔:329;卧龙生 03. 翠袖玉环:402;卧龙生 04. 地狱门:105;卧龙生 05. 飞花逐月:298;卧龙生 06. 飞铃:244;卧龙生 07. 飞燕惊龙:269;卧龙生 08. 风尘侠隐:228;卧龙生 09. 风雨燕归来:198;卧龙生 10. 黑白剑:223;卧龙生 11. 黑白双娇:117;卧龙生 12. 剑气洞彻九重天:299;卧龙生 13. 剑无痕:261;卧龙生 14. 剑仙列传:25;卧龙生 15. 绛雪玄霜:274;卧龙生 16. 金笔点龙记:318;卧龙生 17. 金凤剪:326;卧龙生 18. 金剑雕翎:397;卧龙生 19. 惊鸿一剑震江湖:346;卧龙生 20. 梦幻之刀:106;卧龙生 21. 妙绝天香:28;卧龙生 22. 女捕头:317;卧龙生 23. 飘花令:263;卧龙生 24. 七绝剑:130;卧龙生 25. 七绝剑 II 还情剑:144;卧龙生 26. 情剑无刃:7;卧龙生 27. 琼楼十二曲:181;卧龙生 28. 神州豪侠:261;卧龙生 29. 双凤旗:261;卧龙生 30. 素手劫:374;卧龙生 31. 桃花血令:131;卧龙生 32. 天鹤谱:106;卧龙生 33. 天剑绝刀:321;卧龙生 34. 天龙甲:293;卧龙生 35. 天马霜衣:278;卧龙生 36. 天香飙:246;卧龙生 37. 天涯情侣:113;卧龙生 38. 铁剑玉佩:71;卧龙生 39. 铁苗神剑:358;卧龙生 40. 无名箫:136;卧龙生 41. 无形剑:256;卧龙生 42. 新仙鹤神针:188;卧龙生 43. 血剑丹心:350;卧龙生 44. 烟锁江湖:238;卧龙生 45. 燕子传奇:132;卧龙生 46. 摇花放鹰传:446;卧龙生 47. 一代天骄:353;卧龙生 48. 银月飞霜:132;卧龙生 49. 幽灵四艳:221;卧龙生 50. 玉钗盟:320;卧龙生 51. 玉手点将录:150;卧龙生 52. 袁紫烟:30;卧龙生 53. 岳小钗:308;卧龙生 54. 指剑为媒:57;古龙 01. 白玉老虎:111;古龙 02. 白玉雕龙:24;古龙 03. 碧血洗银枪:81;古龙 04. 边城刀声:50;古龙 05. 边城浪子:62;古龙 06. 彩环曲:76;古龙 07. 残金缺玉:82;古龙 08. 苍穹神剑:131;古龙 09. 楚留香 01 血海飘香:47;古龙 10. 楚留香 02 大沙漠:23;古龙 11. 楚留香 03 画眉鸟:96;古龙 12. 楚留香 04 借尸还魂(鬼恋传奇):41;古龙 13. 楚留香 05 蝙蝠传奇:80;古龙 14. 楚留香 06 桃花传奇:8;古龙 15. 楚留香 07 新月传奇:20;古龙 16. 楚留香 08 午夜兰花:68;古龙 17. 大地飞鹰:56;古龙 18. 大旗英雄传:151;古龙 19. 大人物:37;古龙 20. 多情剑客无情剑:150;古龙 21. 飞刀, 又见飞刀:30;古龙 22. 风铃中的刀声:76;古龙 23. 孤星传:175;古龙 24. 护花铃:193;古龙 25. 欢乐英雄:64;古龙 26. 浣花洗剑录:285;古龙 27. 血鹦鹉:11;古龙 28. 剑·花·烟雨·江南:18;古龙 29. 剑毒梅香:110;古龙 30. 剑客行:187;古龙 31. 剑气书香:27;古龙 32. 剑玄录:160;古龙 33. 九月鹰飞:50;古龙 34. 绝不低头:3;古龙 35. 绝代双骄:249;古龙 36. 猎鹰·赌局:107;古龙 37. 流星·蝴蝶·剑:34;古龙 38. 陆小凤 01 陆小凤前传(金鹏王朝):15;古龙 39. 陆小凤 02 绣花大盗:23;古龙 40. 陆小凤 03 决战前后:26;古龙 41. 陆小凤 04 银钩赌坊:10;古龙 42. 陆小凤 05 幽灵山庄:40;古龙 43. 陆小凤 06 凤舞九天:42;古龙 44. 陆小凤 07 剑神一笑:47;古龙 45. 名剑风流:201;古龙 46. 那一剑的风情:32;古龙 47. 怒剑狂花:102;古龙 48. 飘香剑雨:67;古龙 49. 七杀手:20;古龙 50. 七星龙王:55;古龙 51. 七种武器 01 长生剑:23;古龙 52. 七种武器 02 碧玉刀:22;古龙 53. 七种武器 03 孔雀翎:10;古龙 54. 七种武器 04 多情环:8;古龙 55. 七种武器 05 霸王枪:54;古龙 56. 七种武器 06 离别钩:39;古龙 57. 七种武器 07 拳头:16;古龙 58. 情人箭:197;古龙 59. 三少爷的剑:86;古龙 60. 神君别传:12;古龙 61. 失魂引:96;古龙 62. 天涯·明月·刀:69;古龙 63. 武林外史:179;古龙 64. 湘妃剑:169;古龙 65. 萧十一郎:56;古龙 66. 火并萧十一郎:89;古龙 67. 英雄无泪:54;古龙 68. 游侠录:103;古龙 69. 圆月弯刀:155;古龙 70. 月异星邪:102;李凉 01. 暗器高手:53;李凉 02. 霸枪艳血:34;李凉 03. 百败小赢家:79;李凉 04. 本尊分身:97;李凉 05. 超霸的男人:7;

李凉 06. 超级邪侠:209;李凉 07. 赌棍小狂侠:31;李凉 08. 公孙小刀:69;李凉 09. 红顶记:48;李凉 10. 滑头傻小子:43;李凉 11. 会醉才会赢:189;李凉 12. 活宝小淘气:121;李凉 13. 江湖急救站:118;李凉 14. 江湖双响炮:291;李凉 15. 江湖一担皮:197;李凉 16. 江湖一品郎:186;李凉 17. 惊神关小刀:38;李凉 18. 酒赌小浪子:17;李凉 19. 酒狂任小赌:167;李凉 20. 狂侠南宫鹰:15;李凉 21. 六宝江湖行:72;李凉 22. 矛盾天师:59;李凉 23. 妙贼丁小勾:16;李凉 24. 妙贼丁小勾续集:13;李凉 25. 魔手邪怪:6;李凉 26. 奇神扬小邪续集:98;李凉 27. 奇神杨小邪:102;李凉 28. 忍者龟:77;李凉 29. 神偷小千:17;李凉 30. 淘气世家:81;李凉 31. 天才混混:89;李凉 32. 天才混混外集:27;李凉 33. 天齐大帝:62;李凉 34. 天下第一当:52;李凉 35. 武林嘻游记:180;李凉 36. 小鬼大赢家:29;李凉 37. 小鱼吃大鱼:42;李凉 38. 笑笑江湖:69;李凉 39. 新蜀山剑侠传:20;李凉 40. 新蜀山剑侠传续:1;李凉 41. 杨小邪发威:34;梁羽生 01. 白发魔女传:63;梁羽生 02. 冰川天女传:65;梁羽生 03. 冰河洗剑录:65;梁羽生 04. 冰魄寒光剑:4;梁羽生 05. 草莽龙蛇传:143;梁羽生 06. 大唐游侠传:97;梁羽生 07. 弹指惊雷:77;梁羽生 08. 飞凤潜龙:4;梁羽生 09. 风雷震九洲:135;梁羽生 10. 风云雷电:139;梁羽生 11. 广陵剑:149;梁羽生 12. 瀚海雄风:69;梁羽生 13. 还剑奇情录:8;梁羽生 14. 幻剑灵旗:27;梁羽生 15. 慧剑心魔:102;梁羽生 16. 剑网尘丝:127;梁羽生 17. 江湖三女侠:252;梁羽生 18. 绝塞传烽录:40;梁羽生 19. 狂侠天娇魔女:180;梁羽生 20. 联剑风云录:126;梁羽生 21. 梁羽生传奇:74;梁羽生 22. 龙凤宝钗缘:109;梁羽生 23. 龙虎斗京华:182;梁羽生 24. 鸣镝风云录:134;梁羽生 25. 牧野流星:35;梁羽生 26. 女帝奇英传:49;梁羽生 27. 萍踪侠影录:40;梁羽生 28. 七剑下天山:77;梁羽生 29. 塞外奇侠传:2;梁羽生 30. 散花女侠:48;梁羽生 31. 随笔集:笔不花:4;梁羽生 32. 随笔集:笔花六照:15;梁羽生 34. 武当一剑:94;梁羽生 35. 武林天骄:45;梁羽生 36. 侠骨丹心:103;梁羽生 37. 游剑江湖:275;梁羽生 38. 云海玉弓缘:81;金庸 01 飞狐外传:67;金庸 02 雪山飞狐:29;金庸 03 连城诀:35;金庸 04 天龙八部:126;金庸 05 射雕英雄传:57;金庸 06 白马啸西风:5;金庸 07 鹿鼎记:75;金庸 08 笑傲江湖:199;金庸 09 书剑恩仇录:64;金庸 10 神雕侠侣:81;金庸 11 侠客行:61;金庸 12 倚天屠龙记:145;金庸 13 碧血剑:65;金庸 14 鸳鸯刀:35

4.1.2 风雪

风雪 4.53,卧龙生 01. 镖旗:3;卧龙生 07. 飞燕惊龙:16;卧龙生 08. 风尘侠隐:1;卧龙生 09. 风雨燕归来:1;卧龙生 12. 剑气洞彻九重天:36;卧龙生 15. 绛雪玄霜:4;卧龙生 18. 金剑雕翎:9;卧龙生 19. 惊鸿一剑震江湖:2;卧龙生 22. 女捕头:4;卧龙生 25. 七绝剑 II 还情剑:2;卧龙生 27. 琼楼十二曲:1;卧龙生 31. 桃花血令:2;卧龙生 36. 天香飘:2;卧龙生 38. 铁剑玉佩:7;卧龙生 39. 铁苗神剑:7;卧龙生 42. 新仙鹤神针:15;卧龙生 47. 一代天骄:3;卧龙生 49. 幽灵四艳:1;卧龙生 52. 袁紫烟:6;古龙 07. 残金缺玉:7;古龙 13. 楚留香 05 蝙蝠传奇:1;古龙 14. 楚留香 06 桃花传奇:1;古龙 20. 多情剑客无情剑:4;古龙 23. 孤星传:2;古龙 24. 护花铃:1;古龙 25. 欢乐英雄:3;古龙 29. 剑毒梅香:2;古龙 31. 剑气书香:3;古龙 42. 陆小凤 05 幽灵山庄:1;古龙 46. 那一剑的风情:3;古龙 47. 怒剑狂花:2;古龙 56. 七种武器 06 离别钩:1;古龙 61. 失魂引:14;古龙 63. 武林外史:24;古龙 64. 湘妃剑:5;古龙 67. 英雄无泪:8;李凉 01. 暗器高手:1;李凉 02. 霸枪艳血:5;李凉 03. 百败小赢家:1;李凉 04. 本尊分身:6;李凉 09. 红顶记:1;李凉 11. 会醉才会赢:1;李凉 14. 江湖双响炮:4;李凉 16. 江湖一品郎:1;李凉 19. 酒狂任小赌:32;李凉 31. 天才混混:4;李凉 34. 天下第一当:1;李凉 36. 小鬼大赢家:1;李凉 39. 新蜀山剑侠传:1;李凉 40. 新蜀山剑侠传续:12;梁羽生 01. 白发魔女传:4;梁羽生 02. 冰川天女传:5;梁羽生 03. 冰河洗剑录:1;梁羽生 07. 弹指惊雷:1;梁羽生 09. 风雷震九洲:1;梁羽生 10. 风云雷电:3;梁羽生 12. 瀚海雄风:1;梁羽生 14. 幻剑灵旗:1;梁羽生 20. 联剑风云录:1;梁羽生 21. 梁羽生传奇:1;梁羽生 22. 龙凤宝钗缘:5;梁羽生 25. 牧野流星:5;梁羽生 26. 女帝奇英传:3;

梁羽生 27.萍踪侠影录:3;梁羽生 30.散花女侠:1;梁羽生 37.游剑江湖:1;梁羽生 38.云海玉弓缘:1;金庸 01 飞狐外传:3;金庸 03 连城诀:1;金庸 04 天龙八部:1;金庸 05 射雕英雄传:7;金庸 06 白马啸西风:11;金庸 07 鹿鼎记:1;金庸 10 神雕侠侣:2;金庸 12 倚天屠龙记:1

4.2 拓展任务 1

4.2.1 江湖

同 4.1.1

4.2.2 风雪

同 4.1.2

4.3 拓展任务 2

4.3.1 江湖

卧龙生	江湖	56.28
古龙	江湖	24.79
李凉	江湖	14.51
梁羽生	江湖	14.89
金庸	江湖	4.80

4.3.2 风雪

卧龙生	风雪	128.56
古龙	风雪	86.41
李凉	风雪	74.82
梁羽生	风雪	40.04
金庸	风雪	28.45

参考资料

- [1] <https://blog.csdn.net/jaysonhu/article/details/51672289>
- [2] <https://blog.csdn.net/ming19951224/article/details/81416387>
- [3] <https://blog.csdn.net/gslsxqrj/article/details/82624157>
- [4] <https://stackoverflow.com/questions/10522835/hadoop-java-io-ioexception-mkdirs-failed-to-create-some-path>
- [5] https://blog.csdn.net/jack_bob/article/details/78448758
- [6] <https://blog.csdn.net/dacoolbaby/article/details/84514227>
- [7] <https://zhidao.baidu.com/question/48647849.html>
- [8] https://blog.csdn.net/qq_38526573/article/details/89037097
- [9] <https://blog.csdn.net/nwpuwyk/article/details/27513217>