目录

[一、程序功能描述 1](#_Toc19204749)

[1.提取的异常日志文件 1](#_Toc19204750)

[2.统计的excel表格 1](#_Toc19204751)

[二、生成文件格式说明 1](#_Toc19204752)

[1. 异常日志文件的格式 1](#_Toc19204753)

[2.Excel文件统计格式 2](#_Toc19204754)

[3.使用建议 2](#_Toc19204755)

[二、程序运行参数 2](#_Toc19204756)

[三、实际运行效果 2](#_Toc19204757)

[输出信息说明 3](#_Toc19204758)

[四、关于日志格式的说明 3](#_Toc19204759)

[1. 参数为 bes 时解析的格式 3](#_Toc19204760)

[2.参数为 weblogic 时解析的格式 3](#_Toc19204761)

[3.参数为 weblogic2 时解析的格式 3](#_Toc19204762)

[4.参数为 exception 时解析的格式 3](#_Toc19204763)

[五、扩展自定义格式的方法与编译 3](#_Toc19204764)

[七、缺陷 4](#_Toc19204765)

[八、附录 4](#_Toc19204766)

[1.异常日志文件内容图示（-t 指定非 exception时的情况） 4](#_Toc19204767)

[2.统计Excel文件图（-t 指定非 exception时的情况） 4](#_Toc19204768)

[3.异常日志文件内容图示（-t 为 exception时的情况，默认值） 5](#_Toc19204769)

[4. 统计Excel文件图（-t 为 exception时的情况，默认值） 5](#_Toc19204770)

# 一、程序功能描述

本程序读取指定格式日志的日志文件，并将其中含有异常信息的日志条目提取出来保存在单独的文件，最后将统计不同种类异常的数量，并将统计的结果打印到excel文件中，并存放到特定的目录中。

## 1.提取的异常日志文件

程序会在源日志文件的目录中创建名为<源日志目录/源日志文件名-result/exception-info.log>的新日志文件，新创建的异常日志文件中将只含有源日志文件中包含异常信息的日志条目，同时在每条异常日志的前面添加了其原始的日志文件来源，以及在原始日志文件中的行号，方便查找原始异常日志的上下文信息，以进行后续的异常问题定位。

比如如下列个日志文件，是原始日志

server.log，server1.log

而运行后生成这个文件，是提取后的异常日志。该文件用于记录异常日志在原文件出现的位置。

./server.log-result/exception-info.log

./server1.log-result/exception-info.log

## 2.统计的excel表格

程序会在<源日志目录/源日志文件名-result/>的目录中生成一个excel文件，其文件名为<exception-statistic-yyyyMMdd-HHmmss

.xlsx>，该文件含有按照一定匹配长度计算出来的相同异常日志的数量信息，其结果是对于数量 那一列进行降序排序后的结果，方便查找反复出现的异常日志。

如下所示：

./server.log-result/exception-statistic-20190908-181212.xlsx

# 二、生成文件格式说明

## 1. 异常日志文件的格式

Original log file absolute path: C:\Users\JasonFitch\Desktop\logfetch\0906\guanli\server.log

Original log entry line number : 104

####|2019-08-19 22:22:15.869|SEVERE|org.apache.catalina.core.StandardContext|\_ThreadID=19;\_ThreadName=pool-1-thread-10;\_RequestID=a73e857e-aec7-4a35-aaa7-b7e4857154a0;|WEBTIER101270: Exception starting filter AuthFilter

java.lang.NoClassDefFoundError: org/apache/log4j/Logger

at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)

……ignore some tracestack……

……ignore some tracestack……

……ignore some tracestack……

at java.lang.Thread.run(Thread.java:662)

Caused by: java.lang.ClassNotFoundException: org.apache.log4j.Logger

at org.apache.catalina.loader.WebappClassLoader.loadClass(WebappClassLoader.java:1620)

... 22 more

|####

格式说明：

第一行：该异常日志所属的原始日志文件，用来查找原始的文件。

第二行：该异常日志在原始日志文件中行号，用来查找原始异常位置，并定位其异常上下文信息。

第三行：该异常日志的原始内容，用来查看异常的具体细节,

如果 -t 以exception格式（默认值）启动，则这里的内容是异常的简要信息。

第四行：空行，用来分隔不同的异常日志条目。

## 2.Excel文件统计格式

Exception Key Occur Count

ResTypeMgrService refresh end.CacheStats{hitCount= 134

DictItemService refresh failed:org.springframework 89

DeskSoftServiceImpl.querySysParam Exception: while 84

com.huawei.boss.inventory.common.ReceptionExceptio 13

WEBTIER101418: Method service of Servlet action in 3

格式说明：

Exception Key： 按照参数 -n 指定的匹配长度截取的数据

Occur Count ： 相同Exception Key在原始日志文件中出现的次数。

Excel的输出结果已经按照每种异常的数量进行了降序排序，方便查找大量出现的异常。

## 3.使用建议

可以通过看excel的目录确定当前的信息是哪个日志文件的，然后在相应的异常日志文件中搜索，相应的 Exception Key 部分，然后更具异常日志文件中记录的原始行号找见异常发生的上下文，信息从而定位问题。对于使用exception的分析时，可以通过指定-n 的数值为0来达到只提取异常名的效果。

# 二、程序运行参数

通过直接运行命令可以得到如下帮助信息

java -jar XXX.jar

Missing required option: f

usage: bootstrap.LogAnalyzerMain

-e,--log-encoding <arg> 可选的，解析的日志编码, 默认值： GB2312.

-f,--log-files <arg> 必要的，要解析的日志文件列表,可为目录，不递归。 默认值： 无。

-n,--match-length <arg> 可选的，相同日志匹配长度， 默认值： 50。

-t,--log-type <arg> 可选的，解析的日志格式,目前支持bes,weblogic,weblogic2，

exception具体格式参见<关于日志格式的说明>。 默认值：exception

# 三、实际运行效果

执行如下命令可以看见程序的实际运行效果：

java -jar XXX.jar -f C:\ \0906\\guanli\\server.log

输出如下：

user.dir : C: \target

Analyzer runtime arguments info:

log-type : bes

log-encoding: GB2312

match-length: 100

Target files be processed:

C:\\0906\guanli\server.log

Begin Analyze...

Process C:\\0906\guanli\server.log

Analyze Finished...

Staticics file: C:\\0906\guanli\server.log-result\exception-statistic-20190912-180024.xlsx

## 输出信息说明

1.程序工作目录

2.程序执行参数

3.程序将要处理的文件

4.处理的文件信息，创建的异常日志文件信息，统计结果文件信息

# 四、关于日志格式的说明

可以通过 -t 参数来指定要解析的日志格式，目前不支持混合格式解析，目前可以解析BES和weblogic的如下日志格式，如果有其他的格式需要解析，可以通过增加相应类来填加可以解析的格式，如果提供错误的格式参数，那么程序将会拒绝解析。

## 参数为 bes 时解析的格式

标准bes日志格式模版

####| time | level | module | thread | detail |####

## 2.参数为 weblogic 时解析的格式

weblogic日志格式第一种模版

< time > < level > < module > < code > < info > detail

## 3.参数为 weblogic2 时解析的格式

weblogic日志格式第二种模版

< time > < level > < module > < app > < file > < thread > < source > < blank1 > < blank2 > < timestamp > < code > < detail >

## 4.参数为 exception 时解析的格式

此格式下对日志的格式没有特定的要求。

该模式程序查找日志文件中的所有异常，并提取出异常的名字以及其对应的异常信息，然后统计不同类型的异常数量。

异常信息的长度按照-n参数的匹配长度来确定，如何 -n 传递的是 0 那么程序将只匹配异常的名字。

# 五、扩展自定义格式的方法与编译

需要实现的类有：

entry.AbstractLogRecord

用于结构化存储日志

parser.AbstractLogParser

用于解析日志，按照自定义格式将日志由文本转变为结构化JAVA对象

interceptor.AbstractInterceptor

用于处理日志，比如数量统计，异常提取之类的

analyzer.AbstractAnalyzer

用于整合日志的解析逻辑和处理逻辑

最后在启动类bootstrap.LogAnalyzerMain的bootstrap.LogAnalyzerMain#invokeAnalyzer 方法中添加新增加的解析格式便可。

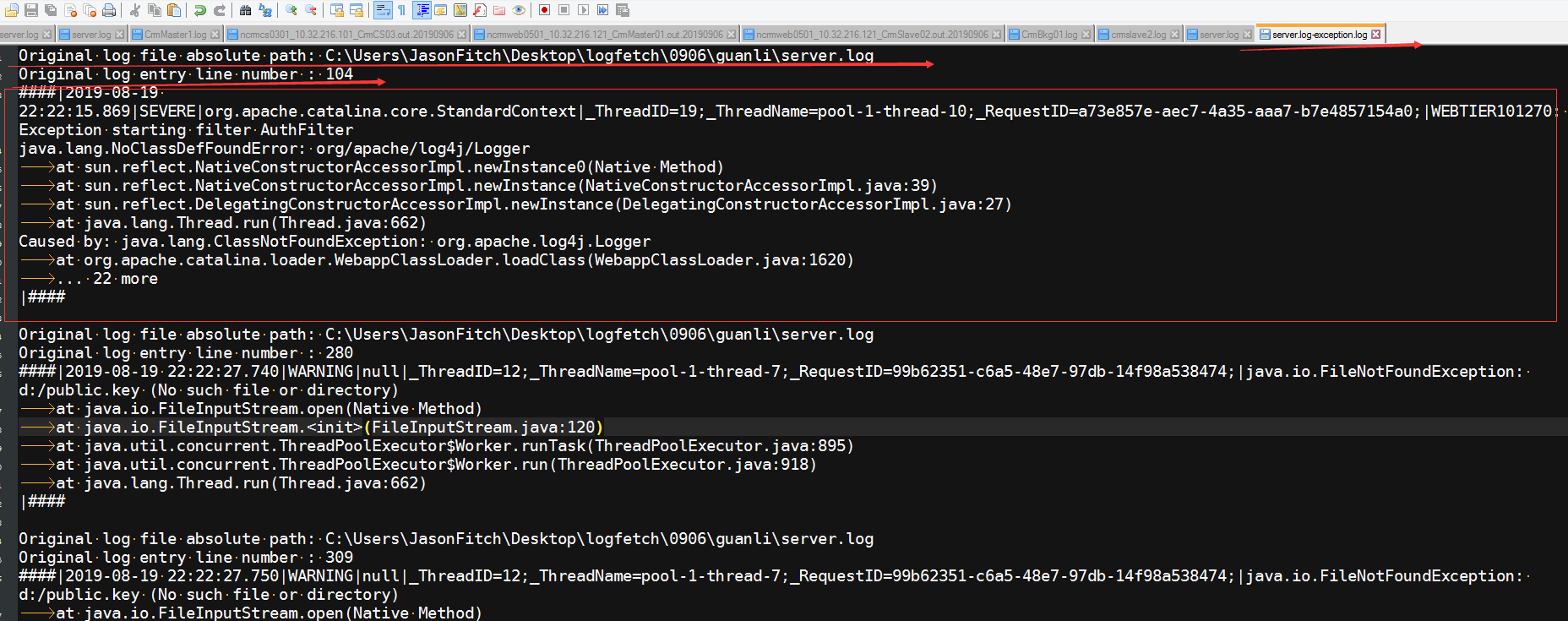
本工程是maven工程,可以直接使用mvn clean install ,生成的jar包都是可执行的。

# 七、缺陷

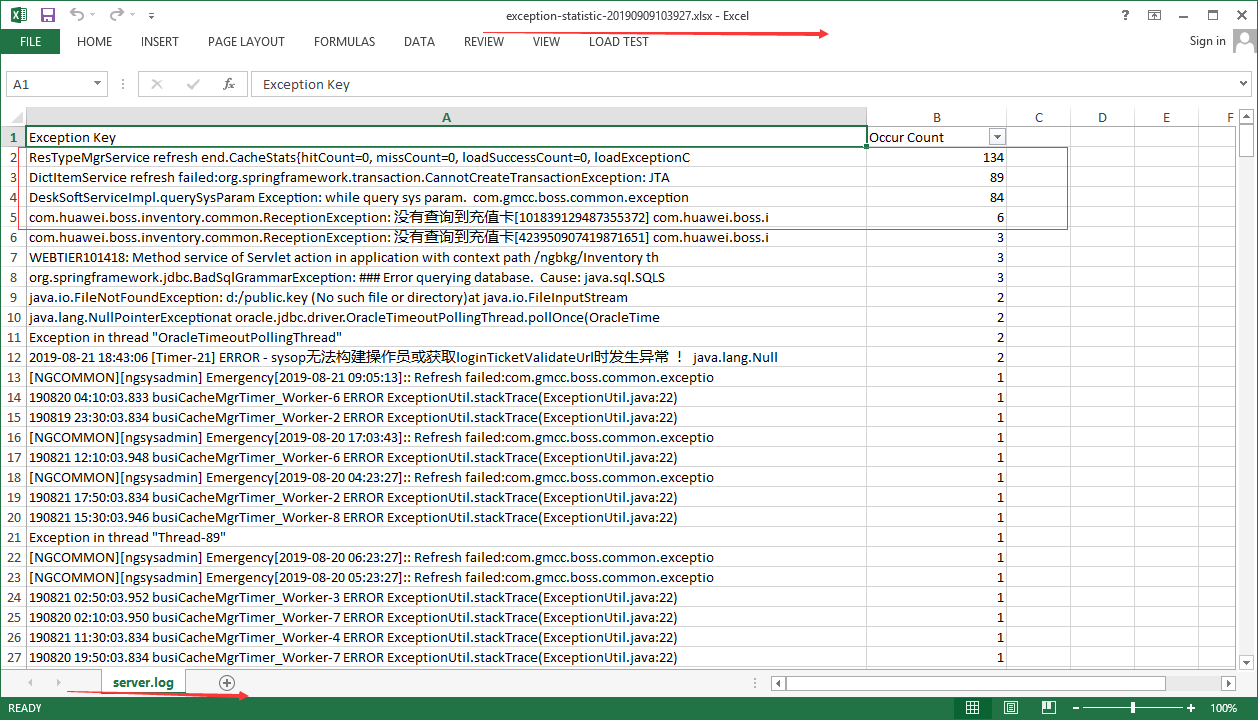
对于日志开头含有时间部分的日志，由于时间不同会被解析成为不同的异常，这个应该可以通过替换掉这些时间部分后在匹配，或者是使用更准确的匹配方式来处理，暂时还没有这样处理。

# 八、附录

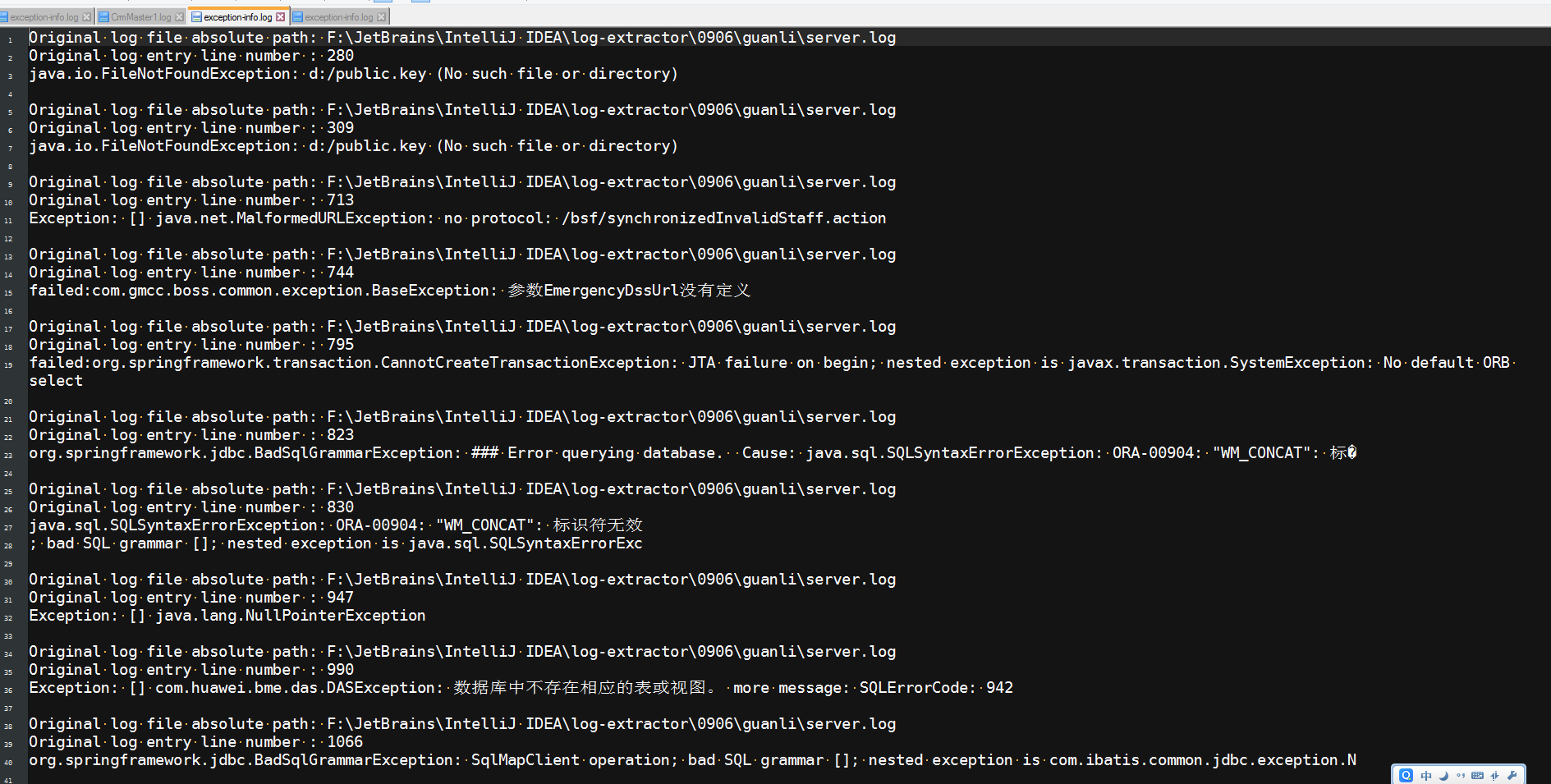
## 1.异常日志文件内容图示（-t 指定非 exception时的情况）



## 2.统计Excel文件图（-t 指定非 exception时的情况）



## 3.异常日志文件内容图示（-t 为 exception时的情况，默认值）



## 4. 统计Excel文件图（-t 为 exception时的情况，默认值）

