第一轮日志解析的总结

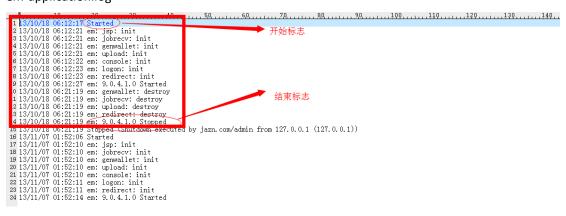
一、 日志的几种类型:

1. 段落式

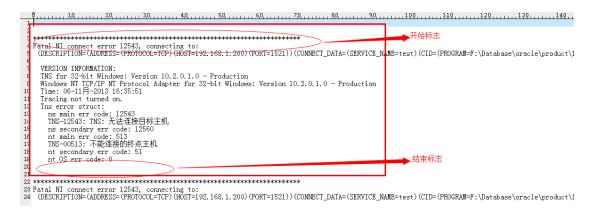
段落式的特点是,有某种特殊的标记开头,中间有多行日志,其中多行日志都是内容相关的,然后按照标记结束(通常都是一个空白行或者是某种结束标记)

log2.log

em-application.log



sqlnet.log



Udev.log



2. 头部固定式

头部固定式,是比较常见的一种日志格式,特点是,一条记录由几个字段构成,其中前面几个字段的形式都是固定的,例如都是日期+线程号+级别+内容组成。其中最后的内容可以是一行也可以是多行。

defmsg.log



tomcat.log

```
1 2012.09.19 18:15:04 Proof Birsester(1472):7 Parse Fatal Error at line 0 column 0: The encoding 'GBK' is not supported.
2 org. mml.sax.9kM arsek_ception. The encoding 'GBK' is not supported.
3 at org. apache. xeroes. Framework. XMLParser, reportError (XMLParser, java:1213)
4 at org. apache. xeroes. Framework. XMLParser, parse (XMLParser, java:1213)
5 at org. apache. xeroes. framework. XMLParser, parse (XMLParser, java:1314)
6 at org. apache. xeroes. framework. XMLParser, parse (XMLParser, java:1314)
7 at org. apache. tomcat. util. digester. Digester. parse the artificationary for intervention of a torg. apache. catalina. startup. ContextConfig. apache. parse the artificationary for intervention of a torg. apache. catalina. startup. ContextConfig. apache. catalina. startup. ContextConfig. apache. (ContextConfig. ispacy)
7 at org. apache. catalina. accord. ContainerDase. addChild(ContainerDase, java:195)
8 at org. apache. catalina. core. ContainerDase. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. core. ContainerDase. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. core. StandardContext. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. core. StandardContext. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. accord. StandardContext. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. accord. StandardContext. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. accord. StandardContext. addChild(ContainerDase, java:1759)
8 at org. apache. catalina. accord. StandardContext. java:1902
9 at org. apache. catalina. accord. StandardContext. java:1903
9 at org. apache. ca
```

3. 表格式

这种形式的日志,类似于一个表格,分为有表头的与无表头的两种。有表头的会在日志 中,包括一个对每个字段的说明。无表头的则只包含数据。

ē.	10	20 20 4	0 50 60	7.0	00 00 100	.,,,,110,,,,,,120,,,,,,,130,,,,,,,140,,
		connections (only server				
		nd-Q Local Address	Foreign Address	State	PID/Program name	
3 tcp	-0	0 0.0.0.0:445	0.0.0.0:*	LISTEN		
4 tcp	0	0 0.0.0.0:993	0. 0. 0. 0:*	LISTEN	_	
5 tcp	0	0 0.0.0.0:995	0.0.0.0:*	LISTEN	_	每个字段的说明
6 tcp	0	0 0.0.0.0:902	0.0.0.0:*	LISTEN	_	每十十段的奶奶
7 tcp	0	0 0.0.0.0:12615	0.0.0.0:*	LISTEN	6005/java	
8 tcp	0	0 0.0.0.0:3306	0. 0. 0. 0:*	LISTEN	-	
9 tcp	0	0 0.0.0.0:6379	0.0.0.0:*	LISTEN	-	
10 tcp	0	0 0.0.0.0:139	0.0.0.0:*	LISTEN	-	
11 tcp	0	0 0.0.0.0:110	0. 0. 0. 0:*	LISTEN	-	
12 tcp	0	0 0.0.0.0:143	0. 0. 0. 0:*	LISTEN	-	
13 tcp	0	0 0.0.0.0:80	0.0.0.0:*	LISTEN	-	
14 tcp	0	0 192.168.122.1:53	0. 0. 0. 0:*	LISTEN	-	
15 tcp	0	0 10.1.1.227:53	0. 0. 0. 0:*	LISTEN	_	
16 tcp	0	0 127. 0. 0. 1:53	0. 0. 0. 0:*	LISTEN	-	
17 tcp	0	0 0.0.0.0:21	0. 0. 0. 0:*	LISTEN	-	
18 tcp	0	0 0.0.0.0:22	0.0.0.0:*	LISTEN	-	
19 tcp	0	0 127.0.0.1:631	0.0.0.0:*	LISTEN	-	
20 tcp	0	0 0.0.0.0:5559	0.0.0.0:*	LISTEN	6006/java	
21 tcp	0	0 0.0.0.0:23	0. 0. 0. 0:*	LISTEN	-	
22 tcp	0	0 0.0.0.0:5560	0. 0. 0. 0:*	LISTEN	6006/java	
23 tcp	0	0 0.0.0.0:5561	0. 0. 0. 0:*	LISTEN	6006/java	
24 tcp	0	0 127. 0. 0. 1:953	0. 0. 0. 0:*	LISTEN	_	

last.log

```
8 monitor pts/2
9 monitor pts/2
10 monitor pts/5
11 monitor pts/5
11 monitor pts/4
12 monitor pts/2
13 monitor pts/2
14 monitor pts/2
15 monitor pts/2
17 monitor pts/3
18 monitor pts/3
18 monitor pts/3
19 monitor pts/2
19 monitor pts/3
20 monitor pts/9
22 monitor pts/9
23 monitor pts/8
23 monitor pts/8
24 monitor pts/6
```

4. 重复式

字段以某种相同的形式重复。 vmenu.log

base_domain.log

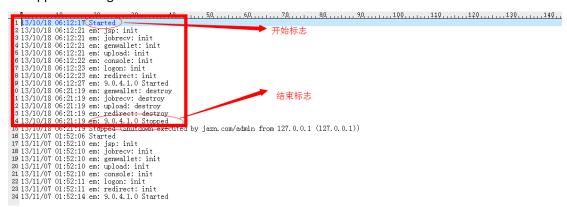
二、目前的解析情况介绍

1. 段落式

能够做到的: 能够在日志中,将段落提取。

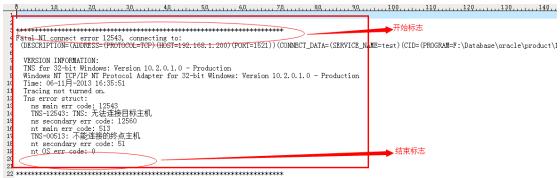
不能做到的: 现阶段无法对提取出的段落做进一步的解析。

em-application.log



No	field1	field2 工格性的	field3	field4
1		### 19	13/10/18 06:21:19	Shutdown executed by jazn.com/admin from 127.0.0.1 (127.0.0.1)
2		13/11/07 01:52:10 em: jsp: init 13/11/07 01:52:10 em: jobrecv: init 13/11/07 01:52:10 em: genwallet: init 13/11/07 01:52:10 em: upload: init 13/11/07 01:52:10 em: console: init 13/11/07 01:52:11 em: logon: init 13/11/07 01:52:11 em: redirect: init 13/11/07 01:52:14 em: 9.0.4.1.0 Started 13/11/07 03:02:22 em: genwallet: destroy 13/11/07 03:02:22 em: jobrecv: destroy 13/11/07 03:02:22 em: upload: destroy 13/11/07 03:02:22 em: redirect: destroy 13/11/07 03:02:22 em: opload: destroy 13/11/07 03:02:22 em: poload: destroy 13/1	13/11/07 03:02:22	JVM termination

sqlnet.log



23 Fatal NI connect error 12543, connecting to:
24 (DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=192.168.1.200)(PORT=1521))(CONNECT_DATA=(SERVICE_NAME=test)(CID=(PROGRAM=F:\Database\oracle\product\]

Fatal NI connect error 12543. connecting to: (DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=192,168,1,200)(PORT=1521))(CONNECT DATA=(SERVICE NAME=test)(CID= (PROGRAM=F:\Database\oracle\product\10.2.0\db_1\jdk\jre\bin\java.exe)(HOST=RAOLIANG)(USER=Administrator))) YERSION INFORMATION: TNS for 32-bit Windows: Version 10.2.0.1.0 - Production Windows NT TCP/IP NT Protocol Adapter for 32-bit Windows: Version 10.2.0.1.0 - Production Time: 06-11月-2013 16:35:51 Tracing not turned on, Tns error structs-ns main err code: 12543 TNS-12543: TNS: 无法连接目标主机 ns secondary err code: 12560 nt main err code: 513 TNS-00513: 不能连接的终点主机 nt secondary err code: 51 nt OS err code: 0

Fatal NI connect error 12543, connecting to: (DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=192.168.1.200)(PORT=1521))(CONNECT_DATA=(SERVICE_NAME=test)(CID=(PROGRAM=F:\Database\oracle\product\10.2.0\db_1\jdk\jre\bin\java.exe)(HOST=RAOLIANG)(USER=Administrator)))) VERSION INFORMATION: TNS for 32-bit Windows: Version 10.2.0.1.0 - Production Windows NT TCP/IP NT Protocol Adapter for 32-bit Windows: Version 10.2.0.1.0 - Production Time: 06-11月-2013 16:36:05 Tracing not turned on. Tns error struct: ns main err code: 12543 TNS-12543: TNS: 无法连接目标主机 ns secondary err code: 12560 nt main err code: 513 TNS-00513: 不能连接的终点主机 nt secondary err code: 51

Fatal NI connect error 12543, connecting to: (DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=192.168.1.200)(PORT=1521))(CONNECT_DATA=(SERVICE_NAME=orcl)(CID (PROGRAM=F:\Database\oracle\product\10.2.0\db_1\jdk\jre\bin\java.exe)(HOST=RAOLIANG)(USER=Administrator)))) VERSION INFORMATION: TNS for 32-bit Windows: Version 10.2.0.1.0 - Production Windows NT TCP/IP NT Protocol Adapter for 32-bit Windows: Version 10.2.0.1.0 - Production Time: 06-11月-2013 16:59:54 Tracing not turned on. Tns error struct: ns main err code: 12543 TNS-12543: TNS: 无法连接目标主机 ns secondary err code: 12560 nt main err code: 513 TNS-00513: 不能连接的终点主机 nt secondary err code: 51 nt OS err code: 0

头部固定式 2.

能够做到的: 能够在日志中, 很准确的识别此记录;

能够准确的提取,头部字段中的值;

能够获得不固定的尾部的文本信息。

不能做到的: 无法对尾部的信息做出进一步的提取与分解。

defmsg.log



No	field1	field2	field3	field4	field5
1	2013-11-29	22:08:10	00001812	00001828	open \\.\DgSafe ok.
2	2013-11-29	22:08:10	00001812	00001828	open device successed.
3	2013-11-29	22:08:10	00001842	00001828	register event successed.
4	2013-11-29	22:17:52	00001812	00001816	open \\.\DgSafe ok.
5	2013-11-29	22:17.52	00001812	00001816	open device successed. 足部内容无法进一步抽
6	2013-11-29	22:17:52	0001812	00001816	register event successed.
7	2013-11-29	22:29:27 前面固定等	00001816解析	00001888	open \\.\DgSafe ok.
8	2013-11-29	22:29:27	00001816	00001888	open device successed.
9	2013-11-29	22:29:27	00001816	00001888	register event successed.

tomcat.log





表格式 3.

能够做到的: 能够很好的解析记录。

包含的 BUG: 在左对齐与右对齐相邻的情况下,会出现解析错误的 bug;

在记录长度小于期望的长度(表头的长度),会出现空指针异常。

Last.log

Ö	. 1.0 2.0	30	4.0	. 5.0	6.0	7.0	8.0	9.0	100	110	120	130 140
1 monitor		10. 1. 1. 112	Thu Mar 2			logged in						
2 monitor	pts/5	10.1.1.112	Thu Mar 2	7 09:14	- 09:16	(00:02)						
3 monitor	pts/3	10. 1. 1. 155	Thu Mar 2	27 07:09	still	logged in						
4 monitor	pts/2	10. 1. 1. 244	Thu Mar 2			logged in						
5 monitor		10. 1. 1. 119	Wed Mar 2			(04:15)						
6 monitor		10. 1. 1. 159	Wed Mar 2			(07:26)						
		10. 1. 1. 155	Wed Mar 2			(10:12)						
8 monitor	pts/2	whosyourdaddy.lo				(10:11)						
9 monitor	pts/6	10. 1. 1. 159	Tue Mar 2			(05:59)						
10 monitor		whosyourdaddy.lo			- 13:06	(09:31)						
11 monitor		whosyourdaddy.lo				(08:11)						
12 monitor		10.1.1.125	Tue Mar 2			(08:00)						
13 monitor		10.1.1.156	Tue Mar 2									
14 monitor	pts/5	10. 1. 1. 155	Mon Mar 2									
15 monitor		10. 1. 1. 116	Mon Mar 2			(04:31)						
16 monitor		10.1.1.156	Mon Mar 2									
17 monitor		10.1.1.155	Mon Mar 2			(09:56)						
18 monitor	pts/2	10.1.1.156	Mon Mar 2 Fri Mar 2			(02:28) (00:27)						
19 monitor	pts/3	10.1.1.244	Fri Mar 2			(00:27)						
20 monitor	pts/10	10.1.1.156	Fri Mar 2			(02:18)						
21 monitor 22 monitor	pts/9	10. 1. 1. 155 10. 1. 1. 155	Fri Mar 2			(02:16)						
23 monitor		10. 1. 1. 155	Fri Mar 2			(04:11)						
24 monitor		10. 1. 1. 155	Fri Mar 2			(00:00)						
24 moni (Or	p (5/ 0	10.1.1.100	iii mar a	.1 05.15	03.13	(00.00)						

	field1	field2	field3	field4	field5
1	monitor	pts/5	10.1.1.112	Thu Mar 27 09:35 stil	logged in
2	monitor	pts/5	10.1.1.112	Thu Mar 27 09:14 - 09:1	(00:02)
3	monitor	pts/3	10.1.1.155	Thu Mar 27 07:09 stil	logged in
4	monitor	pts/2	10.1.1.244	Thu Mar 27 07:03 stil	logged in
5	monitor	pts/5	10.1.1.119	Wed Mar 26 07:22 - 11:3	(04:15)
6	monitor	pts/4	10.1.1.159	Wed Mar 26 04:07 - 11:3	(07:26)
7	monitor	pts/3	10.1.1.155	Wed Mar 26 03:32 - 13:4	(10:12)
8	monitor	pts/2	whosyourdaddy.l	Wed Mar 26 03:32 - 13:4	(10:11)
9	monitor	pts/6	10.1.1.159	Tue Mar 25 05:37 - 11:3	(05:59)
10	monitor	pts/5	whosyourdaddy.l	Tue Mar 25 03:35 - 13:0	(09:31)
11	monitor	pts/4	whosyourdaddy.l	Tue Mar 25 03:35 - 11:4	(08:11)
12	monitor	pts/3	10.1.1.125	Tue Mar 25 03:33 - 11:3	(08:00)
13	monitor	pts/2	10.1.1.156	Tue Mar 25 03:19 - 11:4	(08:22)
14	monitor	pts/5	10.1.1.155	Mon Mar 24 09:27 - 13:4	(04:19)
15	monitor	pts/4	10.1.1.116	Mon Mar 24 07:16 - 11:4	(04:31)

重复式 4.

能够做到的:每个字段的模式完全一样(例如:都是相同的分割符分割),可以很好的解析, 可以提取出每个字段的值。

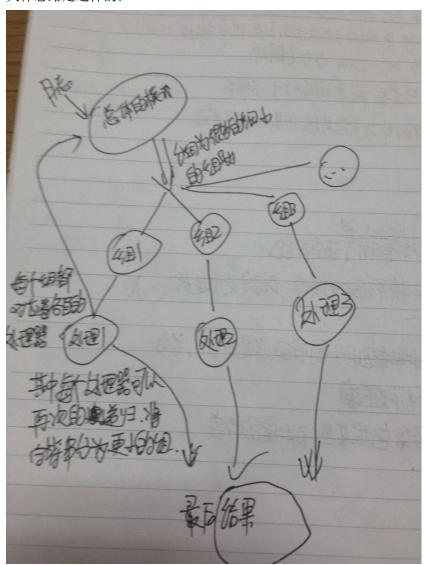
不能做到的: 在此模式中, 只要稍微有点点变化, 都会出现不能匹配的情况; 此模式本质上还是通过正则表达式去匹配的,配置起来非常繁琐,不太容易理 解。

base_domain.log

field1	field2	field3	field4	field5	field6	field7	field8	field9	field10	field11	field12
2012- 3-1下 午02时 58分27 秒 CST	Notice	Security	wolf- haiyi	AdminServer	[ACTIVE] ExecuteThread: '0' for queue: 'weblogic.kernel.Default (self-tuning)'	WLS Kernel			1330585107274	BEA- 090082	Security initializing using security realm myrealm.
2012- 3-1下 午02时 58分28 秒 CST	Notice	WebLogicServer	wolf- haiyi	AdminServer	main	WLS Kernel			1330585108870	BEA- 000365	Server state changed to STANDBY
2012- 3-1 下 午02时 58分28 秒 CST	Notice	WebLogicServer	wolf- haiyi	AdminServer	main	WLS Kernel			1330585108870	BEA- 000365	Server state changed to STARTING

三、下个阶段的目标

1. 在原有的基础上,增加对每个字段进行单独处理的功能。 大体思路是这样的:



- 2. 解决掉表格解析器中两个 bug。
- 3. 优化重复式的配置。