



we\_izheng的专栏

目录视图

摘要视图

RSS 订阅

个人资料



卡卡罗特Z

- 访问：20615次
- 积分：394
- 等级：

BLOG > 2
- 排名：千里之外
  - 原创：19篇 转载：0篇
  - 译文：0篇 评论：13条

文章搜索

文章存档

- 2015年05月(2)
- 2015年01月(2)
- 2014年11月(2)
- 2014年10月(2)
- 2014年08月(3)

展开

阅读排行

- openwrt划分多个vlan(3593)
- 一个修改配置文件的linux shell script(1741)
- 论搜索引擎solr与MongoDB的整合(1732)
- #墙裂推荐Boost regex# C, C++11, Boost三种regex库性能比较(1665)

专家坐阵, Javascript实战分享

微信开发学习路线高级篇上线

免费公开课平台正式上线啦

有奖征文: 云服务器使用初体验

论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎

2015-05-22 00:51

1107人阅读

评论(3)

收藏

举报

solrmongodbjsonmongo-connector嵌套对象

目录(?)	[+]
目录(?)	[+]

有一些东西记录下来，不久之后就会再次用到。

1. 使复杂的嵌套对象从MongoDB更新到Solr

最近使用mongo-connector将MongoDB与Solr打通，对于简单的json，更新是没问题了。这次我们遇到的问题是，如果json中某一个field的值是数组类型，或者有嵌套了一个对象，那该怎么整。比如，我们在MongoDB中插入这么个货：

```
[javascript]
01.  {
02.      "_id": "555df36ec6cd08ea807a4324",
03.      "name": "小米手机",
04.      "comments": [
05.          {
06.              "text": "手机是正品",
07.          },
08.          {
09.              "text": "物流真他妈快",
10.          }
11.      ]
12.  }
```

```
[javascript]
01.  {
02.      "_id": "555df36ec6cd08ea807a4324",
03.      "name": "小米手机",
04.      "comments": [
05.          {
06.              "text": "手机是正品",
07.          },
08.          {
09.              "text": "物流真他妈快",
10.          }
11.      ]
12.  }
```

交叉编译libpcap到openwrt ar71xx(1513)

libpcap, libnet, libnids交叉编译到xilinx上(1428)

基于ssh协议免密码的openwrt环境下数据传输及远程指令执行(1176)

论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎(1100)

cubieboard2 minicom通过USB-TTL线登录(1072)

apache 配置 python cgi(902)

评论排行

#墙裂推荐Boost regex# C, C++11, Boost三种regex库性能比较(4)

[解决]libnids接收不到tcpreply打入的流量数据包(4)

论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎(3)

论搜索引擎solr与MongoDB的整合(1)

一个修改配置文件的linux shell script(1)

基于ssh协议免密码的openwrt环境下数据传输及远程指令执行(0)

基于x86体系结构openwrt上的libmysqlclient交叉编译(0)

关于Boost, C Regex对短目标字符串正则匹配的性能分析(0)

Linux 添加用户及相关配置(0)

vim 重新打开无法回到上次关闭位置(0)

推荐文章

\* 美团Android DEX自动拆包及动态加载简介

\* Android实战技巧之四十四: Hello.Native!

\* 如何面试Python后端工程师?

\* Android自定义控件之仿汽车之家下拉刷新

\* 【Android】 音乐播放器之service服务设计

\* 【FastDev4Android框架开发】 Android MVP开发模式详解(十九)

最新评论

论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎shawzida: 学习了, 可以在博客中引用这篇文章吗? 最近也在使用solr作字符串数组元素的搜索。感觉数组中不同类型...

#墙裂推荐Boost regex# C,

这东西能同步到Solr吗，我手心真是一把汗一把汗啊。经过多次尝试，非常不幸，轰轰烈烈的失败了。

无奈，去我们伟大的官方说明看一看吧，<https://github.com/10gen-labs/mongo-connector/wiki/Usage%20with%20Solr>，其中有一段：Key Names and Document Flattening，尼玛前几天还没有这段说明好不，真是坑爹啊。算了，不管怎样，看起来这东西针对上面那种json还是有办法的，按照说明，上面那段json应该转换成以下形式：

```
[javascript]
01.  {
02.    "_id": "555df36ec6cd08ea807a4324",
03.    "name": "小米手机",
04.    "comments.0.text": "手机是正品",
05.    "comments.1.text": "物流真他妈快"
06.  }
```

```
[javascript]
01.  {
02.    "_id": "555df36ec6cd08ea807a4324",
03.    "name": "小米手机",
04.    "comments.0.text": "手机是正品",
05.    "comments.1.text": "物流真他妈快"
06.  }
```

然后提交给Solr。并且，我们从官方文档中看不出来任何需要配置的地方，难道，默认安装好，这些东西就能行？

Oh, No~~~~

这是为什么呢，我们看看官方对schema.xml的说明，原来关键点在这，大概意思就是mongo-connector会读取这个配置文件，在将数据提交给Solr之前，会将数据中没有在schema.xml中声明的field去掉。

哦，原来如此，应该是形如"comments.0.text"这样的field没有被schemal.xml声明，所以被去掉了，在Solr中见不到。那么我们对症下药，既然没有声明这个域，我们声明不就好了，在schema.xml中加入：

```
[plain]
01.  <field name="comments.0.text" type="text_mmseg4j_complex" indexed="true" stored="true"/>
```

```
[plain]
01.  <field name="comments.0.text" type="text_mmseg4j_complex" indexed="true" stored="true"/>
```

schema.xml的路径参考上一篇文章。重启Solr，保持mongo-connector开启，重新往MongoDB中插入上面那段json，哇，ok，我们可以在Solr前端看到"comments.0.text"了，伟大的革命征程终于迈出了第一步！

成功之后，你可能还想显示"comments.1.text"，如法炮制。那么，问题来了，如果我数组中有多个对象，无限多，难道我要把所有的可能field都声明一遍？

这个问题问的真是极好的！我们不得不说schema.xml是个非常重要的东西！

这个时候，我们要用到在schemal.xml中的另外一种东西，dynamicField。这东西看起来可以通配多个名称，按照已有的示例。那么，按照我们的需求，我们这样添加：

```
[plain]
01.  <dynamicField name="comments*" type="text_mmseg4j_complex" indexed="true" stored="true"/>
```

```
[plain]
01.  <dynamicField name="comments*" type="text_mmseg4j_complex" indexed="true" stored="true"/>
```

这个意思很明显了，就是声明了所有通配comments\*的域，这样我们就可以把所有的comments\*一

- C++11, Boost三种regex库性能比较  
卡卡罗特Z: @CharlesSimonyi:好的, 多谢建议!
- #墙裂推荐Boost regex# C, C++11, Boost三种regex库性能比较  
CharlesSimonyi: 用VC编译器时记得编译成release再测试, 我把你写的C++11 regex那段代码复制制到VS20...
- 论搜索引擎solr与MongoDB的整合  
a642339199: mongod --replSet myDevReplSet这一步骤的时候出错replSet init...
- 论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎  
卡卡罗特Z: @sinat\_28713877: 这个connector确实需要多多研究一下
- 论mongo-connector如何将MongoDB中的json数组和嵌套对象更新至Solr引擎  
sinat\_28713877: 恩 这样可以 我是修改了 solr\_doc\_manager所有修改和增加的数据都要clean\_doc...
- #墙裂推荐Boost regex# C, C++11, Boost三种regex库性能比较  
ahoLic: VS测试的时候, 别用debug啊, 用release试试~
- #墙裂推荐Boost regex# C, C++11, Boost三种regex库性能比较  
115456456456: boost 1.58 不论能不能匹配都是7ms,c++11的regex库不匹配时14s多,太可怕了
- [解决]libnids接收不到tcp replay打入的流量数据包  
Forrest-Gump: @we\_izheng:我使用libnids的时候, 有的pcap文件可以识别, 有的不可以, 但是对比文件...
- [解决]libnids接收不到tcp replay打入的流量数据包  
卡卡罗特Z: @cgz372619:能大概说下问题吗

如果想进一步探究mongo-connector怎样做到上面这些的, 可以参考python文件:

```
[plain] 01. /usr/local/lib/python2.7/dist-packages/mongo_connector/doc_managers/solr_doc_manager.py
```

```
[plain] 01. /usr/local/lib/python2.7/dist-packages/mongo_connector/doc_managers/solr_doc_manager.py
```

到这, 我们就完成了更新这一伟大使命。

## 2. 在Solr中对嵌套对象及数组进行Query

按照1中所述, 我们目前可以在Solr前端看到更新的字段, 并且, 这些字段已经被Solr建立索引, 那么我们该如何查询他们呢, 比如我们想用关键字“正品”在所有的形如“comments\*.text”的字段中命中到“手机是正品”这个值。令人遗憾的是, Solr提供给我们值的各种查询手段, 却不给我们指定目标field的各种手段, 我们只是形如“comments\*.text”, 但事实上, 我们并不能在Solr中这样指定要搜索的域。

这时, 不得不再次说一句, schema.xml真特么重要了, 我们用到了里面又一个东西copyField。这个东西可以看看schema.xml中的注释, 大概意思就是建立索引时source的值可以加入到dest, 反过来查询dest也就同时能对source进行查询, 关键是多个source可以copy到一个dest啊, 这不就正合我意嘛。说了这么多, 只需要我们在schema.xml中加入一句:

```
[plain] 01. <copyField source="comments*" dest="text"/>
```

```
[plain] 01. <copyField source="comments*" dest="text"/>
```

这个dest的field "text"一定要设置成multiValue="true"哦, 不然mongo-connector会报错的, 想来应该也是吧, 这么多source到一个dest, 别问我是怎么知道的, 这个确实是我试出来的, 中间过程就不多说了。

按照上面所说, 这个时候我们对text域进行查找, 就会去coments\*域查找, 设置Solr的df为text, q为关键字, 猛击吧, 骚年, 你会得到想要的东西。

好了, 码了这么多, 关键点就是这么多了, 坑太多, 小心小心呐~~

版权声明: 本文为博主原创文章, 未经博主允许不得转载。

[上一篇](#) [论搜索引擎solr与MongoDB的整合](#)

顶

0

踩

0

### 猜你在找

- Python自动化开发基础 装饰器-异常处理-面向对象编程
- Python 零基础到实战

python安装工具pip、easy\_install  
Python 项目实战\_BBS day14  
Python 项目实战\_cmdb Day15  
Python编程基础视频教程(第五季)  
从此不求人:自主研发一套PHP前端开发框架  
Android入门实战教程  
零基础学Java系列从入门到精通  
【深入理解计算机网络】专家解读计算机网络原理视频讲



查看评论

暂无评论

发表评论

用户名: zhaoyifei1

评论内容:

提交

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

公司简介 | 招贤纳士 | 广告服务 | 银行汇款帐号 | 联系方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

网站客服 杂志客服 微博客服 webmaster@csdn.net 400-600-2320 | 北京创新乐知信息技术有限公司 版权所有 | 江苏乐知网络技术有限公司 提供商务支持

京 ICP 证 070598 号 | Copyright © 1999-2014, CSDN.NET, All Rights Reserved



公司简介 | 招贤纳士 | 广告服务 | 银行汇款帐号 | 联系方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

网站客服 杂志客服 微博客服 webmaster@csdn.net 400-600-2320 | 北京创新乐知信息技术有限公司 版权所有 | 江苏乐知网络技术有限公司 提供商务支持

京 ICP 证 070598 号 | Copyright © 1999-2014, CSDN.NET, All Rights Reserved

