



中国互联网无线化峰会
—暨阿里云开发者大会
CHINA MOBILE INTERNET SUMMIT 2011

2011.10.24
中国·杭州

主办方: Alibaba Group

阿里云

立即报名

论坛首页 → 综合技术版 → mysql → MySQL数据库中文全文检索解决方案

全部 Database Haskell Erlang FP Linux 数据结构和算法 mysql oracle DB2 SQLServer PostgreSQL MacOSX Unix 编程综合 OS

发表 回复

最成熟稳定甘特图控件,支持Java和.Net

浏览 147 次

主题: MySQL数据库中文全文检索解决方案 精华帖 (0) :: 良好帖 (0) :: 新手帖 (0) :: 隐藏帖 (0)

作者

正文

minioa

等级: 初级会员

性别:

文章: 2

积分: 40

来自: 天津

我现在离线

发表时间: 2011-09-30

< >

 猎头职位: [北京: ITeye网站诚聘网站编辑](#)

在MySQL建立英文的全文索引不是难事，但是要做好中文的全文索引需要做很多工作。

MySQL默认不支持中文索引，因为中文不能像英文那样通过空格来准确的判断单词，而需要通过语义来判断，这就需要中文切词工具支持，由于版权问题，MySQL一直没有提供中文支持，虽然MySQL发展的很快。

那么我们只能曲线救国了。

场景模拟，在OA系统中添加对新闻的全文搜索支持。

技术实现，对新闻表的标题和正文字段建立Fulltext索引，然后通过MATCH的方式检索。

虽然我们可以通过like或regexp的方式检索，但是和MATCH相比，前者的效率要远远低于后者，前者需要1秒钟，通过全文索引匹配的后者可能只需要0.002秒钟，这就是效率。

```
SELECT count(*) FROM t where body like '%开源%'
SELECT count(*) FROM t where body regexp '开源'
SELECT count(*) FROM t where body like binary('%开源%')
SELECT body FROM t where match(body) against('开源' in boolean mode)
```

问题难点，英文容易匹配，以为有空格，但是中文一般是一个个的句子，不好判断。

联想思考，要是将中文能够在词语之间加入空格，那么就可以很容易匹配到。经测试，这种方式是可行的。

现在的问题，变成，如何将一个篇幅的中文汉字提取出有效的关键词，然后各关键词用空格连起来，存入一个字段，然后对这个字段建立Fulltext索引。

思路就是这样的，下面的工作就是如何提取关键词，也就是常说的中文切词或分词。

中文切词方案请参照

Lucene全文检索1[url] [http://www.minioa.net/viewtopic.php?f=18&t=264\[url\]](http://www.minioa.net/viewtopic.php?f=18&t=264[url])

Lucene中文检索2[url] [http://www.minioa.net/viewtopic.php?f=18&t=266\[url\]](http://www.minioa.net/viewtopic.php?f=18&t=266[url])

声明：ITeye文章版权属于作者，受法律保护。没有作者书面许可不得转载。

推荐链接

[现在都去万网注册域名了](#)

相关文章:

- JavaEye3.0开发手记之三 — 狮身人面
- 利用thinking sphinx实现全文检索
- 全文检索中近义词、关联词的解决方案

推荐群组: [JAVA 3T](#)
[更多相关推荐](#)

注域名，重服务，15年域名服务经验，选万网，选安全！

www.Net.cn

Google 提供的广告

- 见证又一个准百万富翁的诞生！
- 20-30万急聘多名天才Java/MTA软件工程师
- 3G培训就业月薪平均7K+，不3K就业不花一分钱！

 **Google >> 2011路演**
>> 指导企业创新营销
>> 谷歌专家现场解答企业营销难题

更有机会获
¥3500元
大礼包

[马上了解 >>](#)

[返回顶楼](#)

[主页](#) [资料](#) [短信](#) [留言](#) [关注](#)

hellojinjie

等级: 初级会员



性别: 

文章: 148

积分: 30

来自: 杭州

 我现在离线

发表时间: 2011-09-30

说是解决方案，根本就没有给出解决方案。。。。。

[返回顶楼](#)

[主页](#) [资料](#) [短信](#) [留言](#) [关注](#) [回帖地址](#)

 0

 0

请登录投票

minioa

等级: 初级会员



性别: 

文章: 2

积分: 40

来自: 天津

 我现在离线

发表时间: 2011-10-01

可能我写的不够详细

[返回顶楼](#)

[主页](#) [资料](#) [短信](#) [留言](#) [关注](#) [回帖地址](#)

 0

 0

请登录投票

 [发表回复](#)

[论坛首页](#) → [综合技术版](#) → [mysql](#)

跳转论坛:

- [☐☐: ThoughtWorks☐☐Architect](#)
- [☐☐: Scilearn China☐☐Web ☐☐☐☐☐☐☐](#)

[广告服务](#) | [ITeye黑板报](#) | [联系我们](#) | [友情链接](#)

© 2003-2011 ITeye.com. [[京ICP证110151号](#) 京公网安备110105010620]