

SEO

SEO (Search Engine Optimization): 汉译为搜索引擎优化。搜索引擎优化是一种利用搜索引擎的搜索规则来提高目前网站在有关搜索引擎内的自然排名的方式。SEO的目的理解是: 为网站提供生态式的自我营销解决方案, 让网站在行业内占据领先地位, 从而获得品牌收益; SEO包含站外SEO和站内SEO两方面; SEO是指为了从搜索引擎中获得更多的免费流量, 从网站结构、内容建设方案、用户互动传播、页面等角度进行合理规划, 使网站更适合搜索引擎的索引原则的行为; 使网站更适合搜索引擎的索引原则又被称为对搜索引擎优化, 对搜索引擎优化不仅能够提高SEO的效果, 还会使搜索引擎中显示的网站相关信息对用户来说更具有吸引力。

[【详】](#)

主流的全文索引工具有: [Lucene](#), Sphinx, [Solr](#), [ElasticSearch](#)。其中Solr和Elastic Search都是基于Lucene的。Sphinx不是apache的项目, 如果你想把Sphinx放到某个商业性的项目中, 你就得买个商业许可证。

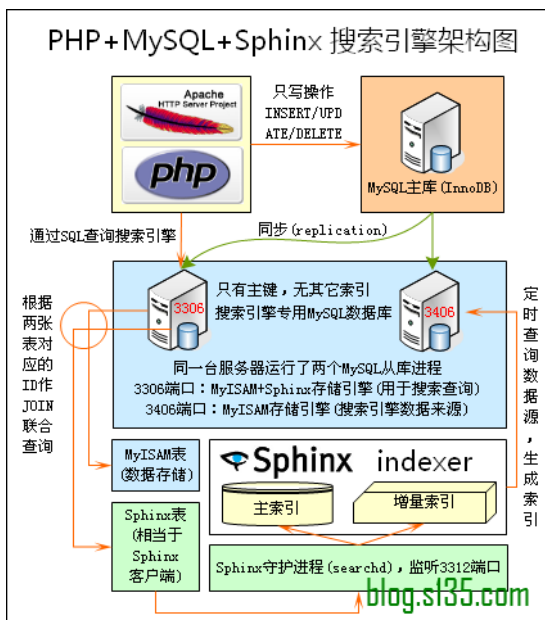
[Sphinx](#)是一款基于SQL的高性能全文检索引擎, Sphinx的性能在众多全文检索引擎中也是数一数二的, 利用Sphinx, 我们可以完成比数据库本身更专业的搜索功能, 而且可以有很多针对性的性能优化。

Sphinx

Sphinx的特点

- 快速创建索引: 3分钟左右即可创建近100万条记录的索引, 并且采用了增量索引的方式, 重建索引非常迅速。
- 闪电般的检索速度: 尽管是1千万条的大数据量, 查询数据的速度也在毫秒级以上, 2-4G的文本量中平均查询速度不到0.1秒。
- 为很多脚本语言设计了检索API, 如PHP,Python,Perl,Ruby等, 因此你可以在大部分编程应用中很方便地调用Sphinx的相关接口。
- 为MySQL设计了一个存储引擎插件, 因此如果你在MySQL上使用Sphinx, 那简直就方便到家了。
- 支持分布式搜索, 可以横向扩展系统性能。

PHP+MySQL+Sphinx 搜索引擎架构图



在MySQL中安装Sphinx

Sphinx在MySQL上安装有两种方式:

- 第一种方式是采用API调用, 我们可以使用PHP,Python,Perl,Ruby等编程语言的API函数进行查询, 这种方式不必重新编译MySQL, 模块间改动比较少, 相对灵活。
- 第二种需要重新编译MySQL, 将Sphinx以插件的方式编译到MySQL中去, 这种方式对程序改动比较少, 仅仅需要改动SQL语句即可, 但前提是你的MySQL版本必须在5.1以上。

下面是第一种安装方式:

```
#下载最新稳定版
wget http://www.sphinxsearch.com/downloads/sphinx-0.9.9.tar.gz
tar xzvf sphinx-0.9.9.tar.gz
cd sphinx-0.9.9
./configure --prefix=/usr/local/sphinx/ --with-mysql --enable-id64
make
make install
```

Sphinx中文分词插件Coreseek安装

注：coreseek的安装教程来自[这里](#)，下面是详细过程：

安装升级autoconf

因为coreseek需要autoconf 2.64以上版本，因此需要升级autoconf，否则会报错。

从<http://download.chinaunix.net/download.php?id=29328&ResourceID=648>下载autoconf-2.64.tar.bz2，安装方法如下：

```
tar -jxvf autoconf-2.64.tar.bz2
cd autoconf-2.64
./configure
make
make install
```

下载coreseek

新版本的coreseek将词典和sphinx源程序放在了一个包中，因此只需要下载coreseek包就可以了。

```
wget http://www.wapm.cn/uploads/csft/3.2/coreseek-3.2.14.tar.gz
```

安装mmseg(coreseek所使用的词典)

```
tar xzvf coreseek-3.2.14.tar.gz
cd mmseg-3.2.14
./bootstrap #输出的warning信息可以忽略，如果出现error则需要解决
./configure --prefix=/usr/local/mmseg3
make && make install
cd ..
```

安装coreseek(sphinx)

```
cd csft-3.2.14
sh buildconf.sh #输出的warning信息可以忽略，如果出现error则需要解决
./configure --prefix=/usr/local/coreseek --without-unixodbc --with-mmseg --with-mmseg-
includes=/usr/local/mmseg3/include/mmseg/ --with-mmseg-libs=/usr/local/mmseg3/lib/ --with-mysql
make && make install
cd ..
```

测试mmseg分词和coreseek搜索

备注：需要预先设置好字符集为zh_CN.UTF-8，确保正确显示中文，我的系统字符集为en_US.UTF-8也是可以的。

```
cd testpack
cat var/test/test.xml #此时应该正确显示中文
/usr/local/mmseg3/bin/mmseg -d /usr/local/mmseg3/etc var/test/test.xml
/usr/local/coreseek/bin/indexer -c etc/csft.conf --all
/usr/local/coreseek/bin/search -c etc/csft.conf 网络搜索
此时正确的应该返回
words:
```

1. '网络': 1 documents, 1 hits
2. '搜索': 2 documents, 5 hits

生成 mmseg词库及配置文件

新版本的已经自动生成。

总结

Sphinx作为一款高性能的SQL全文检索引擎，值得开发者继续关注，尤其是它多语言的API支持，使开发者可以更方便地应用。

