1. Coreseek几条常用命令

简历索引：/usr/local/coreseek/bin/indexer –c 索引文件

启动 ：/usr/local/coreseek/bin/searched –c 索引文件

测试 ：/usr/local/coreseek/bin/search –c 索引文件 keywords

主索引和增量索引的执行任务：

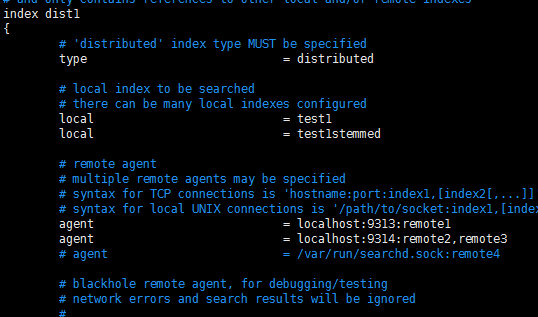
Crontab -1 定是执行

访问量少的情况下我们执行主索引：00 03 \*\*\*/usr/local/coreseek/init/main.sh

每五分钟执行一次增量索引：\*/5\*\*\*\*/usr/local/coreseek/init/delta.sh

分布式索引：

在dist里面配置：



1. 下载及安装

首先要进行中文环境监测：locale

**##以下为核心项，locale为zh\_CN.UTF-8，就可以正常显示和输入中文；locale设置功能由操作系统自身支持，BSD/Linux均可开启；该功能，不是coreseek提供的！！！**

**##如果不能正常显示中文，则后面的中文测试部分无法正常进行，但不会影响coreseek的实际功能；**

**LANG=zh\_CN.UTF-8**

**LC\_ALL="zh\_CN.UTF-8"**

注：如果不是zh\_CN.UTF-8请百度“linux下用locale命令设置语言环境”。

接下来：

1.下载及解压

1. $ wget http://www.coreseek.cn/uploads/csft/3.2/coreseek-3.2.14.tar.gz
2. $ 或者 http://www.coreseek.cn/uploads/csft/4.0/coreseek-4.0.1-beta.tar.gz
3. $ 或者 http://www.coreseek.cn/uploads/csft/4.0/coreseek-4.1-beta.tar.gz
4. $ tar xzvf coreseek-3.2.14.tar.gz 或者 coreseek-4.0.1-beta.tar.gz 或者 coreseek-4.1-beta.tar.gz
5. $ cd coreseek-3.2.14 或者 coreseek-4.0.1-beta 或者 coreseek-4.1-beta

2. ##安装coreseek开发的mmseg，为coreseek提供中文分词功能：

cd mmseg-3.2.14

##ubuntu环境下，需要使用ACLOCAL\_FLAGS="-I /usr/share/aclocal" ./bootstrap

./bootstrap

./configure --prefix=/usr/local/mmseg3

make && make install

接下来进行中文分词测试：

##中文分词测试，如果显示不正常，请检查当前环境下的locale和UTF-8中文字符显示设置

$ /usr/local/mmseg3/bin/mmseg -d /usr/local/mmseg3/etc src/t1.txt

中文/x 分/x 词/x 测试/x

中国人/x 上海市/x

Word Splite took: 1 ms.

3.安装coreseek：

cd csft-3.2.14

##执行configure，进行编译配置：

sh buildconf.sh

./configure

--prefix=/usr/local/coreseek --without-python --without-unixodbc --with-mmseg--with-mmseg-includes=/usr/local/mmseg3/include/mmseg/ --with-mmseg-libs=/usr/local/mmseg3/lib/ --without-mysql

##coreseek4.x/sphinx1.x，请将--without-mysql更改为--with-mysql

make && make install

接下来进行测试：

##配置测试，测试是否可以正确运行

$ /usr/local/coreseek/bin/indexer -c /usr/local/coreseek/etc/sphinx-min.conf.dist

##以下为正常测试时的提示信息：

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file '/usr/local/coreseek/etc/sphinx-min.conf.dist'...

total 0 reads, 0.000 sec, 0.0 kb/call avg, 0.0 msec/call avg

total 0 writes, 0.000 sec, 0.0 kb/call avg, 0.0 msec/call avg

##

##csft-4.0版显示：ERROR: nothing to do.

二、提供数据源支持：

1.我这里只给出mysql和xml数据源支持的说明了：

##系统需要安装相关的基础依赖库，否则无法支持对应的功能

##freebsd6/7/8:

$ pkg\_add -r mysql50-client libxml2 expat

##centos5.4/5.5: fedora12/13 : rhel5.5

$ yum install mysql-devel libxml2-devel expat-devel

##debian5 : ubuntu9/10:

$ apt-get install mysql-client libmysqlclient15-dev libxml2-dev libexpat1-dev

##gentoo10:

##安装环境自带了libmysqlclient libxml2 libexpat

##opensuse10/11:

##yast安装 libmysqlclient-devel libxml2 libxml2-devel libexpat-devel

##slackware13.0/13.1:

$ slapt-get --install mysql libxml2 expat

##archlinux-2010：

$ pacman -S libmysqlclient libxml2 expat

2.重新编译安装coreseek，以支持mysql数据源和xml数据源

$ cd csft-3.2.14

$ make clean

$ ./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

三、全文检索测试：

$ cd testpack

$ /usr/local/coreseek/bin/indexer -c etc/csft.conf

##以下为正常情况下的提示信息：

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

total 0 reads, 0.000 sec, 0.0 kb/call avg, 0.0 msec/call avg

total 0 writes, 0.000 sec, 0.0 kb/call avg, 0.0 msec/call avg

##

##csft-4.0版显示：ERROR: nothing to do.

##

$ /usr/local/coreseek/bin/indexer -c etc/csft.conf --all

##以下为正常索引全部数据时的提示信息：（csft-4.0版类似）

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

indexing index 'xml'...

collected 3 docs, 0.0 MB

sorted 0.0 Mhits, 100.0% done

total 3 docs, 7585 bytes

total 0.075 sec, 101043 bytes/sec, 39.96 docs/sec

total 2 reads, 0.000 sec, 5.6 kb/call avg, 0.0 msec/call avg

total 7 writes, 0.000 sec, 3.9 kb/call avg, 0.0 msec/call avg

$ /usr/local/coreseek/bin/indexer -c etc/csft.conf xml

##以下为正常索引指定数据时的提示信息：（csft-4.0版类似）

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

indexing index 'xml'...

collected 3 docs, 0.0 MB

sorted 0.0 Mhits, 100.0% done

total 3 docs, 7585 bytes

total 0.069 sec, 109614 bytes/sec, 43.35 docs/sec

total 2 reads, 0.000 sec, 5.6 kb/call avg, 0.0 msec/call avg

total 7 writes, 0.000 sec, 3.9 kb/call avg, 0.0 msec/call avg

$ /usr/local/coreseek/bin/search -c etc/csft.conf

##以下为正常测试搜索时的提示信息：（csft-4.0版类似）

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

index 'xml': query '': returned 3 matches of 3 total in 0.093 sec

displaying matches:

1. document=1, weight=1, published=Thu Apr 1 22:20:07 2010, author\_id=1

2. document=2, weight=1, published=Thu Apr 1 23:25:48 2010, author\_id=1

3. document=3, weight=1, published=Thu Apr 1 12:01:00 2010, author\_id=2

words:

$ /usr/local/coreseek/bin/search -c etc/csft.conf -a Twittter和Opera都提供了搜索服务

##以下为正常测试搜索关键词时的提示信息：（csft-4.0版类似）

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

index 'xml': query 'Twittter和Opera都提供了搜索服务 ': returned 3 matches of 3 total in 0.038 sec

displaying matches:

1. document=3, weight=24, published=Thu Apr 1 12:01:00 2010, author\_id=2

2. document=1, weight=4, published=Thu Apr 1 22:20:07 2010, author\_id=1

3. document=2, weight=3, published=Thu Apr 1 23:25:48 2010, author\_id=1

words:

1. 'twittter': 1 documents, 3 hits

2. '和': 3 documents, 15 hits

3. 'opera': 1 documents, 25 hits

4. '都': 2 documents, 4 hits

5. '提供': 0 documents, 0 hits

6. '了': 3 documents, 18 hits

7. '搜索': 2 documents, 5 hits

8. '服务': 1 documents, 1 hits

$ /usr/local/coreseek/bin/searchd -c etc/csft.conf

##以下为正常开启搜索服务时的提示信息：（csft-4.0版类似）

Coreseek Fulltext 3.2 [ Sphinx 0.9.9-release (r2117)]

Copyright (c) 2007-2010,

Beijing Choice Software Technologies Inc (http://www.coreseek.com)

using config file 'etc/csft.conf'...

listening on all interfaces, port=9312

##如要停止搜索服务，请使用/usr/local/coreseek/bin/searchd -c etc/csft.conf --stop

##如要已启动服务，要更新索引，请使用/usr/local/coreseek/bin/indexer -c etc/csft.conf --all --rotate

##然后，请参考csft-3.2.14下api目录中的相关文件，使用PHP、Python、Ruby、Java来测试搜索服务；也可以前往< ahref="/products-install/step\_by\_step/">搜索服务建立三步曲，查看第三步使用PHP测试。

四、php测试全文检索：

<?php

//注意文件的编码格式需要保存为为UTF-8格式

require("sphinxapi.php");// 在coreseek3.2.14/testpack/api 下

$cl = new SphinxClient ();

//以下设置用于返回数组形式的结果

$cl->SetArrayResult ( true );

/\*

//ID的过滤

$cl->SetIDRange(3,4);

//sql\_attr\_uint等类型的属性字段，需要使用setFilter过滤，类似SQL的WHERE group\_id=2

$cl->setFilter('group\_id',array(2));

$cl->SetFilterRange('group\_id2',6,8);

\*/

//取从头开始的前20条数据，0,20类似SQl语句的LIMIT 0,20

$cl->SetLimits(0,20);

//在做索引时，没有进行 sql\_attr\_类型 设置的字段，可以作为“搜索字符串”，进行全文搜索

$res = $cl->Query ('test', "\*" ); //"\*"表示在所有索引里面同时搜索，"索引名称（例如test或者test,test2）"则表示搜索指定的

//$res=cl->Query( ‘测试’ , "\*");//测试中文检索

//print\_r($res);

//如果需要搜索指定全文字段的内容，可以使用扩展匹配模式：

//$cl->SetMatchMode(SPH\_MATCH\_EXTENDED);

//$res=cl->Query( '@title (测试) @content ('网络')' , "\*");

//print\_r($res['matches']);

print\_r($res);

//print\_r($cl->GetLastError());

//print\_r($cl->GetLastWarning());

echo '</pre>';

?>