参考：<http://blog.csdn.net/daiyan_csdn/article/details/74536599>

**Lucene和solr学习目录**

**几个概念先行说明一下：**  
Lucene: 是一个索引与搜索类库，而不是完整的程序。

Solr：是一个高性能，采用Java5开发，基于Lucene的一个独立的企业级搜索应用服务器，它对外提供类似于Web-service的API接口。

Nutch：是一个由**[Java](http://lib.csdn.net/base/java" \o "Java 知识库" \t "_blank)**实现的，刚刚诞生开放源代码(open-source)的web[**搜索引擎**](http://lib.csdn.net/base/searchengine)。

[**Hadoop**](http://lib.csdn.net/base/hadoop)：  实现了一个分布式文件系统（**[hadoop](http://lib.csdn.net/base/hadoop" \o "Hadoop知识库" \t "_blank)** Distributed File System），简称HDFS。于 2005 年秋天作为 Lucene的子项目 Nutch的一部分正式引入。它受到最先由 Google Lab 开发的 Map/Reduce 和 Google File System(GFS) 的启发。2006 年 3 月份，Map/Reduce 和 Nutch Distributed File System (NDFS) 分别被纳入称为 [**Hadoop**](http://lib.csdn.net/base/hadoop) 的项目中。

**1. Lucene**

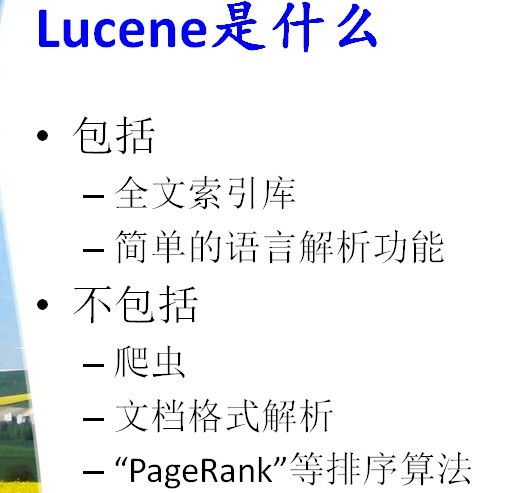
**1.1. Lucene是什么**

　　Lucene是apache软件基金会4 jakarta项目组的一个子项目，是一个开放源代码的全文检索引擎工具包，即它不是一个完整的全文检索引擎，而是一个全文检索引擎的**[架构](http://lib.csdn.net/base/architecture" \o "大型网站架构知识库" \t "_blank)**，提供了完整的查询引擎和索引引擎，部分文本分析引擎（英文与德文两种西方语言）。

　　Lucene的目的是为软件开发人员提供一个简单易用的**工具包**，以方便的在目标系统中实现全文检索的功能，或者是以此为基础建立起完整的全文检索引擎。

**Lucene: 是一个索引与搜索类库，而不是完整的程序。**

**使用Lucene的方式主要有二种：一是自己编写程序，调用类库；二是使用第三方基于Lucene编写的程序，如下面介绍的Solr等。**



　　更多详情，可查看百度百科的说明： <http://baike.baidu.com/view/371811.htm>

　　相关中文文档，可以查看Solr中文网：[http://www.solrcn.com/](http://www.solrcn.com/" \t "_blank)

**1.2. 资源下载**

[10\_Lucene 原理与代码分析完整版.pdf](http://download.csdn.net/detail/liuweitoo/4724696)

[11\_LuceneInAction(第2版)\_Cn.pdf](http://download.csdn.net/detail/liuweitoo/4698568)  （**可作为学习的重点**，本人学习笔记请点击这里查看）

　　12\_Lucene+nutch[**搜索引擎**](http://lib.csdn.net/base/searchengine)开发.pdf  ([Part1](http://download.csdn.net/detail/liuweitoo/4698701) 　　[Part2](http://download.csdn.net/detail/liuweitoo/4698762))

　　13\_开发自己的搜索引擎——Lucene+Heritrix(第2版) ([Part1](http://download.csdn.net/detail/liuweitoo/4698614)　　[Part2](http://download.csdn.net/detail/liuweitoo/4698621))

[13\_开发自己的搜索引擎——Lucene+Heritrix(第2版)\_随书光盘](http://download.csdn.net/detail/liuweitoo/4698680)

[Hadoop权威指南(第2版).pdf](http://download.csdn.net/detail/liuweitoo/4724802)

**1.3. Lucene学习**

[LuceneInAction(第2版)学习笔记——目录](http://blog.csdn.net/liuweitoo/article/details/8137381)

[Lucene索引文件结构速览](http://blog.csdn.net/liuweitoo/article/details/8149184)

**1.4.  Lucene索引查看工具Luke**

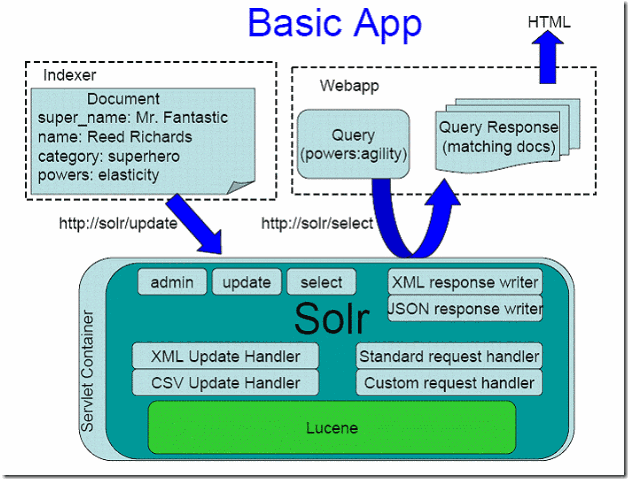
　　可以点击这里下载：[http://code.google.com/p/luke/](http://code.google.com/p/luke/" \t "_blank)

**2. Solr**

**2.1. Solr是什么**

　　Apache Solr 是一个开源的搜索服务器，Solr 使用 [**Java**](http://lib.csdn.net/base/java) 语言开发，主要基于 HTTP 和 Apache Lucene 实现。定制 Solr 索引的实现方法很简单，用 POST 方法向 Solr 服务器发送一个描述所有 Field 及其内容的 XML 文档就可以了（me:类似es构建索引，http post json文档）。定制搜索的时候只需要发送 HTTP GET 请求即可，然后对 Solr 返回的信息进行重新布局，以产生利于用户理解的页面内容布局。

**Solr原理图**



**Solr用户**



**2.2. 资源下载**

　　21\_Lucene\_Solr\_搜索引擎解密.ppt

　　22\_开源企业搜索引擎Solr的应用教程.doc　（[快速上手特别推荐，请点击这里下载](http://download.csdn.net/detail/liuweitoo/4698840)）

　　23\_Solr技术文档.pdf

　　24\_solr3.1 cookbook.pdf

　　25\_Apache+Solr+3+Enterprise+Search+Server.pdf

　　26\_apache-solr-beyond-the-box.pdf

　　27\_Packtpub.Solr.1.4.Enterprise.Search.Server.Aug.2009.pdf

　　上述文档打包下载： [请点击这里](http://download.csdn.net/detail/liuweitoo/4698880)

**2.3. Solr原理**

**2.3.1. Solr服务原理**

　　Solr对外提供标准的http接口来实现对数据的索引的增加、删除、修改、查询。

　　在Solr中，用户通过向部署在servlet 容器中的Solr Web应用程序发送 HTTP 请求来启动索引和搜索。

　　Solr接受请求，确定要使用的适当SolrRequestHandler，然后处理请求。通过 HTTP 以同样的方式返回响应。  
　　默认配置返回Solr的标准 XML 响应，也可以配置Solr的备用响应格式。

**2.3.2. 索引原理**

 　　可以向Solr索引servlet传递四个不同的索引请求：  
　　1) add/update允许向Solr添加文档或更新文档。直到提交后才能搜索到这些添加和更新。  
　　2) commit 告诉Solr，应该使上次提交以来所做的所有更改都可以搜索到。  
　　3) optimize 重构 Lucene 的文件以改进搜索性能。索引完成后执行一下优化通常比较好。如果更新比较频繁，则应该在使用率较低的时候安排优化。一个索引无需优化也可以正常地运行。优化是一个耗时较多的过程。  
　　4) delete 可以通过 id 或查询来指定。按 id 删除将删除具有指定 id 的文档；按查询删除将删除查询返回的所有文档。

**2.3.3. 搜索原理**

 　　要实现搜索则只需要调用搜索接口select发送HTTP GET。示例：  
<http://192.168.10.85:18080/solr/select?indent=on&version=2.2&q>= solr&start=0&rows=10&fl=\*%2Cscore&qt=standard&wt=standard  
　　注：多核心时为这个地址[http://192.168.10.85:18080/solr/core0/select/](http://192.168.10.85:18080/solr/core0/select/" \t "_blank)

**2.4. Solr源码结构**

　　请查看另一篇《[Solr源码结构](http://blog.csdn.net/liuweitoo/article/details/8129135#t1)》，主要内容如下：

**2.4.1. 目录结构说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8129135#t0)

**2.4.2. Solr home说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8129135#t8)

**2.4.3. Solr的包说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8129135#t13)

**2.5. Solr安装配置**

**2.5.1. Linux下的安装与配置Solr**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8129512)

**2.5.2. Windows下安装与配置Solr(tomcat7.0)**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8129632)

　　附：《[Windows下安装JDK并配置环境变量](http://blog.csdn.net/liuweitoo/article/details/8131781)》

**2.5.3. 多核配置**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8132708)

**2.5.4. Solr配置文件说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8132981)

**2.5.5. Solr配置文件schema.xml和solrconfig.xml分析**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8137124)

**2.5.6. Solr中文分词配置(IKAnalyzer)**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8133532)

**2.5.7. 在Jetty容器运行Solr----标准运行方式**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8135061)

**2.5.8. 导入Solr默认的示例数据**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8135120)

**2.5.9. 为多核Solr实例添加一个支持默认示例数据的核**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8135491)

**2.6. Solr管理与维护**

**2.6.1. Solr的admin页面说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8136073)

**2.6.2. Solr搜索结果说明**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8136742)

**2.6.1. Solr基础理论与维护管理快速上手(含查询参数说明)：(非常重要)**

[点此查看](http://blog.csdn.net/liuweitoo/article/details/8136760)

**2.7.  Solr的Facet**

　　网上某牛人写的非常棒的《[31\_solr+facet学习笔记.docx](http://download.csdn.net/detail/liuweitoo/4724588" \t "_blank)》，值得收藏，本人的Facet入门文档。

　　《[Solr的Facet学习笔记与个人总结](http://blog.csdn.net/liuweitoo/article/details/8141024)》

**2.8.  Solr的客户端**

**2.8.1. SolrJ**

《[利用SolrJ操作solr API完成index操作](http://blog.csdn.net/liuweitoo/article/details/8137023" \t "_blank)》

**2.8.2. SolrNet**

**2.9. Solr4.0**