2.2.1系统架构设计

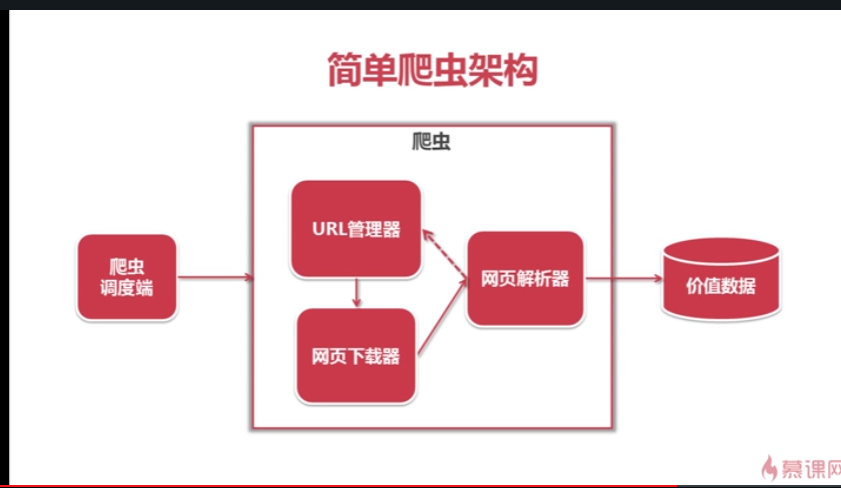
本文所述系统的硬件架构主要包括以下四大部分：

1)门户区：主要负责前台交互。

2)检索区：处理【门户区】发来的检索请求并向【门户区】反馈检索结 果。

3)索引区&分析区：负责接收请求、反馈结果并且负责源文件数据交互。

4)存储区：主要负责数据存储。



[1]县小平. 搜索引擎PageRank算法研究[D].西北大学,2010.

[1]吴淑燕,许涛. PageRank算法的原理简介[J]. 图书情报工作,2003,(02):55-60+51.

1.2.2 MySQL

MySQL是一种关系数据库管理系统，关系数据库将数据保存在不同的表中，而不是将所有数据放在一个大仓库内，这样就增加了速度并提高了灵活性。MySQL所使用的 SQL 语言是用于访问[数据库](http://baike.baidu.com/view/1088.htm)的最常用标准化语言。MySQL 软件采用了双授权政策，分为社区版和商业版，由于其体积小、速度快、总体拥有成本低，尤其是[开放源码](http://baike.baidu.com/view/394804.htm)这一特点，一般中小型网站的开发都选择 MySQL 作为网站数据库。由于其社区版的性能卓越，搭配 [PHP](http://baike.baidu.com/view/99.htm) 和 [Apache](http://baike.baidu.com/view/28283.htm) 可组成良好的开发环境。

MySql与SQL Server相比具有如下优势：

1. 使用 C和 [C++](http://baike.baidu.com/view/824.htm)编写，并使用了多种[编译器](http://baike.baidu.com/view/487018.htm)进行测试，保证了[源代码](http://baike.baidu.com/view/60376.htm)的可移植性。
2. 支持 [AIX](http://baike.baidu.com/view/349664.htm)、[FreeBSD](http://baike.baidu.com/view/21459.htm)、HP-UX、[Linux](http://baike.baidu.com/view/1634.htm)、[Mac OS](http://baike.baidu.com/view/157344.htm)、[Novell](http://baike.baidu.com/view/68826.htm)Netware、[OpenBSD](http://baike.baidu.com/view/337596.htm)、OS/2 Wrap、Solaris、[Windows](http://baike.baidu.com/view/4821.htm)等多种操作系统。
3. 为多种[编程语言](http://baike.baidu.com/view/552871.htm)提供了 [API](http://baike.baidu.com/view/16068.htm)。这些编程语言包括 [C](http://baike.baidu.com/view/10075.htm)、C++、[Python](http://baike.baidu.com/view/21087.htm)、[Java](http://baike.baidu.com/view/29.htm)、[Perl](http://baike.baidu.com/view/46614.htm)、[PHP](http://baike.baidu.com/view/99.htm)、Eiffel、Ruby,.NET和 Tcl 等。
4. 既能够作为一个单独的应用程序应用在[客户端服务器](http://baike.baidu.com/view/3028815.htm)网络环境中，也能够作为一个库而嵌入到其他的软件中。
5. 提供多语言支持，常见的编码如中文的 [GB 2312](http://baike.baidu.com/view/93250.htm)、[BIG5](http://baike.baidu.com/view/98718.htm)，日文的 [Shift\_JIS](http://baike.baidu.com/view/1568180.htm)等都可以用作数据表名和数据列名。
6. 提供 [TCP/IP](http://baike.baidu.com/view/7729.htm)、ODBC 和 [JDBC](http://baike.baidu.com/view/25611.htm)等多种[数据库](http://baike.baidu.com/view/1088.htm)连接途径。
7. 提供用于管理、检查、优化数据库操作的[管理工具](http://baike.baidu.com/view/1251417.htm)。
8. 支持大型的数据库。可以处理拥有上千万条记录的大型数据库。
9. 支持多种[存储引擎](http://baike.baidu.com/view/1238930.htm)。
10. MySQL 是开源的，所以你不需要支付额外的费用。
11. MySQL 使用标准的 [SQL](http://baike.baidu.com/view/34.htm)数据语言形式。
12. MySQL 对 PHP 有很好的支持，[PHP](http://baike.baidu.com/subview/99/5828265.htm)是目前最流行的 Web 开发语言。

1.1编程语言

服务器端语言主要有Java,PHP等。

1.1.1 [Java](http://baike.baidu.com/subview/29/12654100.htm)

[Java](http://baike.baidu.com/subview/29/12654100.htm)是一门[面向对象](http://baike.baidu.com/view/125370.htm)编程语言，不仅吸收了[C++](http://baike.baidu.com/view/824.htm)语言的各种优点，还摒弃了C++里难以理解的[多继承](http://baike.baidu.com/view/1160382.htm)、[指针](http://baike.baidu.com/view/159417.htm)等概念，因此Java语言具有功能强大和简单易用两个特征。Java语言作为静态面向对象编程语言的代表，极好地实现了面向对象理论，允许程序员以优雅的思维方式进行复杂的编程。

1.1.2 PHP

PHP（外文名:PHP: Hypertext Preprocessor，中文名：“[超文本](http://baike.baidu.com/view/156868.htm)[预处理器](http://baike.baidu.com/view/499651.htm)”）是一种通用[开源](http://baike.baidu.com/view/9664.htm)[脚本语言](http://baike.baidu.com/view/76320.htm)。[语法](http://baike.baidu.com/view/135635.htm)吸收了[C语言](http://baike.baidu.com/view/1219.htm)、[Java](http://baike.baidu.com/view/29.htm)和[Perl](http://baike.baidu.com/view/46614.htm)的特点，利于学习，使用[广泛](http://baike.baidu.com/view/344354.htm)，主要适用于[Web](http://baike.baidu.com/view/3912.htm)开发领域。PHP 独特的[语法](http://baike.baidu.com/view/135635.htm)混合了[C](http://baike.baidu.com/view/10075.htm)、[Java](http://baike.baidu.com/view/29.htm)、[Perl](http://baike.baidu.com/view/46614.htm)以及[PHP](http://baike.baidu.com/view/99.htm)自创的语法。它可以比[CGI](http://baike.baidu.com/view/32614.htm)或者[Perl](http://baike.baidu.com/view/46614.htm)更快速地执行[动态网页](http://baike.baidu.com/view/348756.htm)。用PHP做出的[动态页面](http://baike.baidu.com/view/2065821.htm)与其他的[编程语言](http://baike.baidu.com/view/552871.htm)相比，[PHP](http://baike.baidu.com/subview/99/5828265.htm)是将[程序](http://baike.baidu.com/view/17674.htm)嵌入到[HTML](http://baike.baidu.com/view/692.htm)（[标准通用标记语言](http://baike.baidu.com/view/5286041.htm)下的一个应用）文档中去执行，执行效率比完全生成[HTML](http://baike.baidu.com/view/692.htm)标记的[CGI](http://baike.baidu.com/subview/32614/12037322.htm)要高许多；PHP还可以执行[编译](http://baike.baidu.com/view/69568.htm)后代码，编译可以达到[加密](http://baike.baidu.com/view/40927.htm)和[优化](http://baike.baidu.com/view/548.htm)代码运行，使代码运行更快。

Java语言虽然很强大，但是PHP开发语言有如下的优势：

1. 开发源代码：所有的PHP[源代码](http://baike.baidu.com/view/60376.htm)事实上都可以得到。
2. 免费性：和其它技术相比，PHP本身免费且是开源代码。
3. 快捷性：程序开发快，运行快，技术本身学习快。嵌入于[HTML](http://baike.baidu.com/view/692.htm)：因为PHP可以被嵌入于[HTML](http://baike.baidu.com/view/692.htm)语言，它相对于其他语言。[编辑](http://baike.baidu.com/view/28525.htm)简单，实用性强，更适合初学者。
4. 跨平台性强：由于PHP是运行在服务器端的[脚本](http://baike.baidu.com/view/54.htm)，可以运行在[UNIX](http://baike.baidu.com/view/8095.htm)、[LINUX](http://baike.baidu.com/view/1634.htm)、[WINDOWS](http://baike.baidu.com/view/4821.htm)、[Mac OS](http://baike.baidu.com/view/157344.htm)、[Android](http://baike.baidu.com/subview/1241829/9322617.htm)等平台。
5. 效率高：[PHP](http://baike.baidu.com/subview/99/5828265.htm)消耗相当少的[系统资源](http://baike.baidu.com/view/53557.htm)。
6. 图像处理：用PHP动态创建[图像](http://baike.baidu.com/view/42116.htm),PHP图像处理默认使用GD2。且也可以配置为使用image magick进行图像处理。
7. 面向对象：在php4,php5 中，面向对象方面都有了很大的改进，php完全可以用来开发大型商业程序。
8. 专业专注：PHP支持脚本语言为主，同为类C语言。

综上所述，对于中小型的网站而言，选择PHP作为后台语言是更好地选择。

1.2数据库

现在数据库管理系统主要有SQL Server, MySQL等。

1.2.1 SQL Server

SQL Server是Microsoft 公司推出的关系型[数据库管理系统](http://baike.baidu.com/view/68446.htm)。具有使用[方便](http://baike.baidu.com/view/142800.htm)可伸缩性好与相关[软件](http://baike.baidu.com/view/37.htm)集成[程度](http://baike.baidu.com/view/644590.htm)高等[优点](http://baike.baidu.com/view/1465058.htm)，可跨越从[运行](http://baike.baidu.com/view/1026025.htm)Microsoft Windows 98 的膝上型[电脑](http://baike.baidu.com/view/2358.htm)到运行Microsoft Windows 2012 的大型多处理器的服务器等多种平台使用。