



WebSearch Server 1.0

站内检索系统

产品白皮书



北京盈商动力软件开发有限公司

北京盈商动力软件开发有限公司

2005-12-15

在网站中，有大量的需要被反复利用的有价值信息，如何科学高效地去管理这些信息，使它们在最短的时间内最大限度地发挥其价值，是每一个网站面临的问题。

网站全文检索系统就是将用户的网站内容由系统自动地添加到检索引擎（数据库）中，这样，网站的访问者就可以用关键词方式查询网站中任何他所关心的内容，而不需繁琐的人工查找。

传统的信息管理的方式是采用以纸张为载体的信息分类管理，容量小，保存难。随着计算机产业的发展，以计算机存储设备为载体的电子文档随即出现，但是，人们仍然沿用传统的分类管理的方式来管理这些电子文档，或者根本就没有任何管理手段，这样就无法将这些信息有效的管理和利用，造成了信息巨大的浪费。

全文检索的出现，将彻底改变传统的信息处理的模式。全文检索的基本原理就是以各类数据诸如文字、声音、图象等为处理对象，提供按照数据资料的内容而不是外在特征来实现信息的检索。它能提供快捷的数据管理工具和强大的数据查询手段，通过快捷的数据管理工具，能快速帮助人们进行文档资料的整理和管理工作，而通过强大的数据查询手段，使人们能很方便地查到他们想要的任何信息。对于网站比较庞大、网页数目比较多、内容丰富的企业来讲。网站全文检索可以提供有效的查询手段，来增强网站的吸引力，方便访问者快速有效地找到自己感兴趣的信息。全文检索实现了网站的数据库管理模式，使网站由静态响应转变为智能化动态响应，极大地提高了网站的应用水平。

全文检索系统能为您做什么？

- 每天自动将新发布的网页内容增加到全文索引数据库中，每个访问网站的用户可以在第一时间内检索到您发布到网站的信息。
- 将您单位有价值的信息在 Internet/Intranet 网上发布。
- 您单位或个人的站点信息较多，可方便上网查找所需信息者进行站内检索。
- 如果您是 Internet 内容供应商（ICP），可为您的用户提供更方便的信息查询手段。
- 如果您是 Internet 服务提供商（ISP），可帮助您为用户提供更完美的服务，以增强您

的竞争力。

WebSearch 全文检索系统组成模块

- **全文数据库服务器** 核心技术为中英文全文检索，以卓越的性能和多种强大的功能，为各种格式文档的存储、管理和检索提供原动力。
- **站点检索工具** 为 Web 站点提供该站点网页的全文检索功能，由 Web 模块和全文数据库服务器组成。Web 模块为最终用户的使用提供简单易用的操作界面。可自动定期搜索网站，自动根据网站内容更新索引。
- **网页抓取工具** 可通过“智能机器人”工具（在检索系统中不包括该模块）从互联网上成批或有选择的收集用户感兴趣的信息，能够抓取文本、图像、多媒体文件和各种格式化文档。拥有灵活的抓取控制策略，如控制抓取层次等。高效率和高可靠性的抓取，能跳过无效链接。自动化程度高，能自动启动 Spider 搜索功能的时间，以及重复搜索的时间间隔。操作非常简单便捷。
- **与 WebBuilder 内容管理和发布系统的集成** 可方便地按各种设置将数据库中的数据自动生成 HTML 文件，加载到 Web 服务器上，并可按设定的时间每天自动进行，不需人工干预。可自动根据数据库字段的取值生成导航页面和相关静态网页。为构造电子化媒体网站而专门设计，主要完成网上特定站点信息的采集和整理，分为数据采集和数据加工两部分。所有的操作可以通过浏览器完成。根据用户的分类标准把用户的资料按层次分门别类地组织在一起，生成静态索引页面，便于用户按目录浏览的方式定位所需信息。
- **搜索引擎构造工具** 由 GetSite 网页抓取模块、可以与 WebBuilder 内容管理和发布系统进行集成，与全文数据库服务器等集成。该工具可以为门户网站定制他们特有的专业搜索引擎或通用搜索引擎。

WebSearch 全文检索系统突出特点

- **查询速度快** 基于全文检索技术检索网页的信息，比传统的数据库检索技术和直接检索网页的内容要快，降低了对检索服务器的硬件要求并提高了响应速度。

- **智能化高查准率、查全率** 实现了以自然语言处理技术为核心内容的智能中文分词技术，能够满足查全率和查准率的要求。
- **查询手段丰富** 支持各种逻辑运算（如 AND、OR、NOT 等），支持西文检索和中西文混合检索、词频运算和按时间或相关性（重要性）排序，等等。
- **新网页自动入库** 由 WebBuilder 系统生成的网页将自动进入全文索引数据库，无需用户进行任何干预。
- **多平台支持** 支持多种硬件平台（如小型机、工作站、服务器、PC 机等）和多种操作系统（如 Linux、Sun Solaris、Windows /NT/2000 等）。
- **多语言支持** 支持简体中文（GB2312、GBK）、繁体中文（BIG5）、日文、韩文和英文，并能够在不同的编码之间转换。
- **高开放性** 可透明地连接 Oracle、SQL Server、DB2、Informix、Sybase 等主流的关系数据库管理系统。
- **其它优秀特点** 如简便的安装、便捷的用户管理，等等。