

网站信息管理发布系统设计与应用

The Design and Applications of WebSite Content Manage & Publish System

杜义华 张亚 (中国科学院计算机网络信息中心管理服务中心 100864)

摘要:本文介绍一套包含采编发全过程、采用规范化网页生成技术和信息预发布工作模式的网站信息管理发布系统,包括功能设计、运行环境、标准流程、系统组成,并提供根据实际需求在大型 Portal 网站、中型专业网站、小型企业网站及办公平台等中的应用方案。

关键词:信息发布系统 网站内容管理 Content Manage System

目前狭义的网站信息管理系统很多,它们参差不齐、各有侧重,广义的网站信息管理系统相对少见,比较有代表性的有 TRS 系列软件产品^[2]、新浪网站信息管理系统(SinaIMS)^[1],但应用复杂、价格昂贵。多年来我们进行中国科学院网站、中国互联网信息中心网等一系列网站的开发建设,设计和开发有一套完整的网站信息管理和发布系统,介绍如下。

1 发布系统功能设计

网站信息管理系统一般均能进行类别管理、信息管理、模板管理、权限管理、支持生成静态网页,内置 WYSIWYG(所见即所得)编辑器等,在此基础上,我们新设计以下功能和特色:

1.1 提供信息采编发全套自动化管理

提供有从 Spider 新闻信息实时采集系统、信息专题编辑管理平台、网页生成程序、文件交换程序等全套管理。其中稿件来源除由各级编辑在其权限范围内供稿外,还提供有实时信息抓取技术,可以自动跟踪相关网站信息更新情况,及时将可用的信息资源获取和规整加入资源数据库中,辅助信息采编工作。

1.2 采用规范化网页生成技术,提高页面响应速度

提供自动发布、手工发布和定时发布多种页面生成方法,支持 HTML 和 XML + XSL 两种页面生成格式^[3],允许灵活命名静态目录结构。能将对各末级页面和各栏目、专题的首页按不同模板从动态网关程序文件发布为相互链接的 Web 页面,使其在用户访问网页、查看相关链接、查看更多网页时不必每次需要连接数据库各表查询,从而加强信息的重组和利用,提高服务器性能和响应速度。

1.3 采用信息预发布的工作模式,保证网站的严肃、稳定和

安全性

为网站建立发布服务器和预发布服务器,将信息采集、信息抓取、编辑管理、信息重组、专题建立或栏目调整、页面发布更新及制作调试等工作放在预发布服务器上进行,避免网站运行中可能因正在编辑或调试造成的链接有错或信息有误,让日常耗时耗资源的生成静态页面和更新页面超级链接等不影响对外网页的性能和正常运行。同时,发布与预发布服务器相互交换数据和文件,互为备份,增强网站安全性。

1.4 采用动态频道组织技术,能灵活定制各类专题

系统采用动态频道栏目组织技术,频道栏目可以灵活调整和无限分层而不需要改变任何程序或模块,保证平台扩展性和生命力。并可以轻松灵活新建专题,根据需要重组库内信息,快速推出专栏。

1.5 具有丰富实用页面版式功能,支持全文核心主题标引

信息支持图片附件、交叉分类和分页显示;提供标题搜索、全文搜索、图片新闻、热点排行、当前位置、相关文章、推荐、评论、点击情况等实用功能,并可选择性进行全文核心主题标引链接。加入关键词链接为根据整理后的一些专门词汇,将网页中相关字词智能分析后加入超链接,可充分发挥互联网优势,让各信息、栏目、网站相互连通。

1.6 提供强大方便栏目信息管理平台和页面生成系统

编辑系统具有所见即所得的可视化编辑环境,图文混排,有附件管理面板,能指定不同网页模版,提供预览功能,支持 HTML 和文本模式切换、多媒体内容及自动排版、跨栏目发布功能, B/S 结构版编辑系统能远程发布、异地管理, C/S 版中对栏目专题可以实时热调整而不需担心相关信息、图片、授权、版面的错乱。

1.7 提供灵活授权与审批机制、确保信息发布安全

具备完整的角色权限管理,各频道、栏目或子栏目能单独授权到不同角色和用户;支持数据分级审核发布流程,支持信息一审、二审,提供详细编审发日志,确保所发布信息准确性。

2 发布系统流程与组成

2.1 网站发布管理运行应用环境拓扑图与说明

以大中型综合网站为例,运行及应用环境构建如图 1 所示。

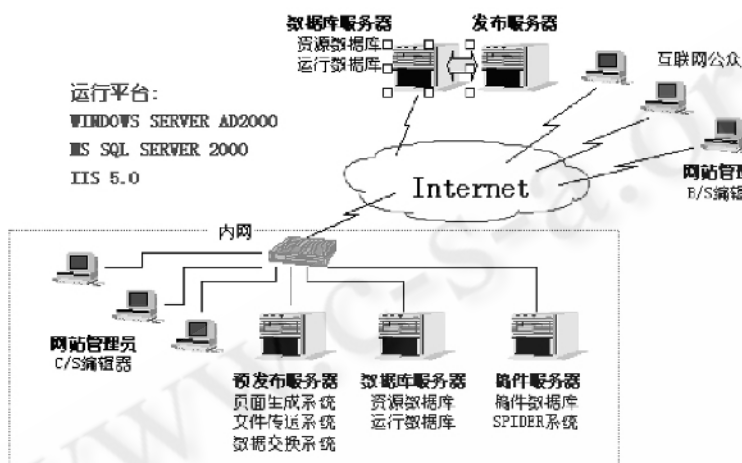


图 1 网站运行应用拓扑图

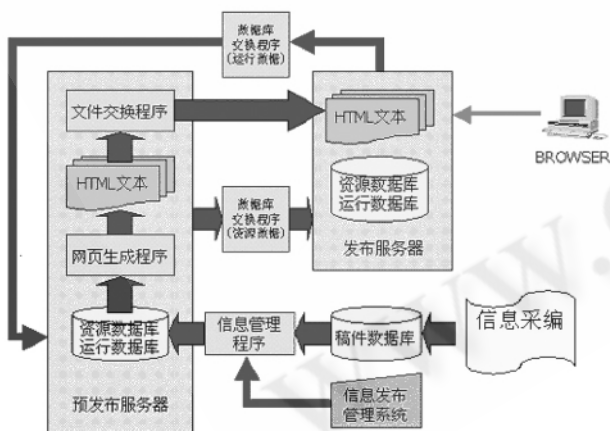


图 2 系统采编发实现流程示意图

(1) 发布服务器节点。网站对公众的发布用服务器,放置于外网。包括 WWW 服务和用于数据检索的 DB 服务器。发布服务器上发布的为标准 HTML 页面,若不需要支持数据检索脚本,可不受运行平台的限制。根据网站规模和服务器处理能力,可将 WWW 服务与 DB 服务合在一台服务器或在

需要时采用服务器集群方式处理。

(2) 预发布服务器节点。预发布服务器与发布服务器相互呼应互为备份,主要承担发布信息的集成工作。预发布服务器也包括 WWW 服务器和数据库服务器,作用包括镜像备份,信息管理,静态页面生成,网页制作调试等,运行环境为 WINDOWS SERVER AD2000 + IIS 5.0,数据库为 MS SQL SERVER 2000,为了安全,一般位于内网。根据网站规模、信息更新频率及页面生成、交换服务程序的配置方式,预发布服务器负担可能较轻或很重,根据需要可将 WWW 服务器和 DB 服务器共用一台或分开。

(3) 稿件服务器节点。稿件服务器用于为预发布中正式数据库服务器提供稿源,方便编辑的操作处理。其资料来源为各下属部门的在线投稿,或从其他系统如政务系统、内部办公网等的定时导入或为利用 SPIDER 技术自动定期抓取的网页信息。

稿件服务器上数据库 MS SQL SERVER 2000,提供有对其他数据库的数据交换接口。一般位于内网。根据网站规模,稿件来源渠道等,稿件服务器功能也可以并入预发布服务器中。

(4) 网站管理员节点。网站管理员实际包括有网站普通供稿人员、编辑人员、审批人员、网页制作改版(技术支持)人员等,承担着网站信息采集和页面调整的运行职能。均能通过 B/S 或 C/S 方式操作控制网站内容和运行,其计算机为普通

PC 机,运行环境为 Windows + IE,位于内网和外网。

(5) 公众用户节点(最终访问者)。

2.2 发布系统标准流程

发布流程按照文档报批的工作流程多级权限进行。低一级权限只能发布到系统中而无法将信息发布到互联网上,只有具备一定权限的用户才可以对系统中的信息做审核并最终发布到网页前台供互联网上的用户浏览。即:

录入信息——一审审核——二审审核——生成 HTML 网页——发布(自动或定时上传)

系统实现流程如图 2 所示。

- (1) 信息采集至稿件服务器;
- (2) 编辑通过平台在稿件服务器选稿、编辑、排版后转入预发布环境的资料数据库,状态为待审核;
- (3) 主编(高一级编辑)对负责频道栏目的信息进行审核,标记为待发布;
- (4) 预发布服务器上的页面生成服务程序监测到有需要发布的信息页面和栏目页面等后自动生成 Html 文件。
- (5) 预发布服务器上的文件交换服务程序检查到有新

生成 Html 页面或新上传的图片附件等后自动上传到发布服务器和在本地做好备份。

(6) 预发布服务器上的信息类数据定期通过数据交换程序更新至发布服务器;发布服务器上记录访问者交互类信息定期通过数据交换程序更新至预发布服务器(数据相互备份)。

2.3 发布系统组成

(1) B/S 版信息管理及发布平台。用于在线投稿,远程管理;提供可视化访问 Word 编辑界面管理信息、图片和附件;能进行频道栏目树查看管理、会员与授权管理、发布模板定制管理等,对信息的审批发表,并根据需要可进行直接生成静态 html 页面或标记为有效。

(2) C/S 版信息编辑与栏目管理。用于快速进行频道栏目及信息图片的查询、增改和调整;分多窗对照操作,各类快捷键的支持;鼠标拖放调整网站栏目间关系、信息分类,全方位栏目组合查询定位和信息检索;专题管理和专题资料再收集,信息图片可视化管理、主题词库生成管理及各类统计与打印。

(3) 信息自动采集和整理程序。通过建立网页资源库、结合 Spider 技术、内容分析技术,引入用户自定义数据项编辑器和替换处理指令编辑器等,提供和定制可视化通用性较强的互联网信息采集系统,能定期自动跟踪相关网站或网页,进行比较分析、抽取、规整入库、分类等从互联网上获取所需从无序到有序化后为网站的稿件来源^[4]。

(4) 页面生成服务程序。为按一定规则将主页中 asp 等网关程序转换为相应的静态页面程序,适用于大型网站的发布应用。能灵活配置发布模板,发布周期,发布内容等;不受 web 服务平台限制,能进行扫描批量发布和个别文件的单独发布;支持各类索引页的自动翻页显示;支持自动加关键词链接等;可以手工操作或作为一个服务驻留自动运行;采用先进先出与优先级干预结合方式的任务序列处理,有较好容错保护和日志记录机制。

(5) 生成关键词链接程序。利用互联网的优势,最大可能将互联网上各类知识点相互连接;可以对不同目录扫描操作;支持输入一段文字效果预览;能对词库表编辑管理,词库可以从栏目表、信息表中生成;提供接口供页面生成程序调用;

(6) 文件交换服务程序。用于预发布与发布服务器间文件同步。为自动监的服务程序,在预发布服务器上实时运行,能将工作目录下新生成的网页拷贝或上传至发布服务器和预发布服务器,然后清空自己。程序具有是否含子目录、是否操作所有类型文件、是否同时清除空目录、是否自动取消只读文件的只读属性等选项。

(7) 数据库交换服务程序。用于数据的相互备份,即预发布服务器上的信息类数据定期通过数据交换程序更新至发布服务器;发布服务器上记录访问者交互类信息定期通过数据交换程序更新至预发布服务器。

3 发布系统三种应用方案

发布系统是一个较广泛的概念,从商业门户网站的新闻系统到内部办公平台、个人的 Weblog(网络上的一种流水记录形式)都可以称作或包含发布系统^[5]。以上系统提供灵活底层框架和丰富应用模块,但具体使用时,需根据访问量、访问对象、信息量、信息更新频率、信息敏感级别、权限控制、版式要求、互动功能需求、硬软件环境、网络环境等实际需求和条件,进行方案选择和环节配置。大致可分为以下三种方案:

3.1 大型网站(Portal 系统)应用方案

(1) 适合类型:信息日更新频繁,专题建设多,信息量大,采编工作量大,信息大多没有密级,访问量大;需要生成静态页面。

(2) 体系结构:同以上的完整介绍,分稿件服务器、预发布服务器、发布服务器等;

(3) 系统组成:各级编辑、主编在局域网内时主要使用 c/S 版信息编辑与栏目管理工具快速管理,在异地可以通过 b/S 版信息编辑器和发布平台远程管理,预发布服务器上自动运行有页面生成服务程序、文件交换服务程序,必要时可以通过 b/S 版信息发布平台手工控制生成页面。其他如信

息自动采集和整理程序、生成关键词链接程序、数据库交换服务程序等则为可选组件。

(4) 应用案例:中国科学院网 <http://www.cas.cn>;

医药博览网 <http://www.100md.com>

3.2 中型网站(专业网站)应用方案

(1) 适合类型:信息更新一般,访问量较大,信息大多没有密级;需要生成静态页面,分发布服务器与预发布服务器;

(2) 体系结构:为大型网站中体系结构的裁剪,稿件服务器与预发布服务器合在一起,网页发布 web 服务器放于外网,其他位于内网或外网均可。编辑访问预发布服务器可通过 C/S 或 B/S 两种方式进行。预发布服务器上生成的文件通过程序自动 FTP 或手工干预 FTP 到网页发布服务器上,对外公布的是网页发布服务器。

(3) 运行环境:外部发布系统基于 NT 或 linux,内部预发布系统基于 NT;Linux 系统下配置 Apache 服务和 MySQL 数据库,WINDOWS 平台配合 IIS 服务和 MS SQL SERVER 2000。

(4) 系统组成:各级编辑、主编主要使用 B/S 版信息编辑器和发布平台,预发布服务器上运行有文件交换服务程序,为最高管理员提供 C/S 版信息编辑与栏目管理工具,其

(下转第 72 页)

他可选组件如信息自动采集和整理程序、生成关键词链接程序、数据库交换服务程序等不必采用。

(5) 应用案例: 中科院互联网信息中心网站 <http://www.cnnic.cn>; 中科院产业化信息网 <http://it.cas.cn>

3.3 小型网站、办公系统应用方案

(1) 适合类型: 访问量较少或局域网内运行, 信息的有一定密级, 需记录页面被点击情况和被哪些用户已阅读等信息, 不一定需要生成 HTML 页面, 如研究所网站、企业网站、政府网站、小型行业网站。

体系结构和运行环境均很简单和单一, 可只需一台服务器, 授权后的信息管理人员使用 B/S 编辑器版本即可, 只为最高管理员提供 C/S 版信息编辑与栏目管理工具。

(2) 应用案例: 中科院研究所内外网站系统 <http://swgk.cashq.ac.cn>

中科院产业局工作网站 <http://chanye.cashq.ac.cn>

4 结束语

目前各类网站生成和信息发布系统很常见, 但真正功能

完整易用的并不多。以上系统在中国科学院“十五”信息化重大项目“科学院主页系统建设”中独立设计开发, 并在众多大小网站和办公系统中广泛应用、日益完善。

参考资料

- 1 新浪网信息发布管理系统, <http://soft-ssi.sina.net/publish.html>。
- 2 北京拓尔思 (TRS) 信息技术有限公司网站, <http://www.trs.com.cn/>。
- 3 杜义华, 内容管理系统中基于 XML 页面发布的实现与应用, 计算机编程技巧与维护, 2004. 7。
- 4 杜义华、及俊川, 通用互联网信息采集系统的设计与开发, 计算机应用研究, 2005. 1。
- 5 车东, 内容管理系统 (CMS) 的设计和选型, <http://www.chedong.com/tech/cms.html>。