

基于 Web2.0 的 Wiki 系统的研究与应用

李学俊, 李龙澍, 徐 怡

(安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039;

安徽大学 计算机学院, 安徽 合肥 230039)

摘 要: 目前互联网 Web2.0 的广泛应用和 Wiki 系统的发展已经成为人们广泛关注的焦点。介绍了新一代互联网 Web2.0 的特征和应用。概要地介绍了 Wiki 的来源、发展历史以及具体应用系统。详细研究了 Wiki 语言的实现机制和 Wiki 系统的实现机制及其功能, 包括 Wiki 页面编辑、版本比较、知识分类、系统管理等。给出了具体应用实例。

关键词: Web2.0; Wiki 语言; Wiki 系统

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2007)07-0250-04

Research and Application of Wiki System Based on Web2.0

LI Xue-jun, LI Long-shu, XU Yi

(Ministry of Education Key Lab. of IC & SP, Anhui University, Hefei 230039, China;

Computer School, Anhui University, Hefei 230039, China)

Abstract: The widely application of Web2.0 and the development of Wiki system have become research focus. Firstly the features and application of the new generation Web2.0 are introduced, then the original and history of Wiki is introduced in brief. Secondly the realization mechanism of Wiki Language, and the implementation mechanism and function of Wiki system are researched in detail. The Wiki's functions include Wiki page edition, version difference, knowledge classification, system management and so on. Finally the application example is given.

Key words: Web2.0; Wiki language; Wiki system

0 引 言

在 Web1.0 时代,网民多是看客,没有信息的控制权,处于被动不平等地位。随着网络技术的飞速发展,尤其是 XML 交换技术的成熟和 P2P 技术的风行,终于为 Web2.0 这场革命拉开了序幕。已经出现了很多 Web2.0 应用,如博客(Blog)、聚合资讯(RSS, Really Simple Syndication)、播客(Podcasting)、社会化书签(Social Bookmark)、个人社会网络(SNS, Social Network Software)、维基 Wiki 等。

作为 Web2.0 的典型应用, Wiki 集中网民的集体智慧来构建网络公共知识库,如全世界最大的 Wiki 项目“维基百科全书”。Wiki 还推广到更宽广的领域,完

成更加多样化的任务,从菜谱的讨论、剧本的创作到芯片的设计、飞机的制造无所不包,只要有发起者,就不缺支持者。因此文中全面深入地研究了 Wiki 系统,包括 Wiki 语言的实现机制和 Wiki 系统的实现机制及其功能,并给出了具体应用实例^[1]。

1 Web2.0

Web2.0 这一概念,由 O'Reilly 媒体公司总裁兼 CEO 提姆·奥莱理提出。Web2.0 目前没有一个统一的定义。这里只给出 Web2.0 的特征描述。Web2.0 是互联网的一次理念和思想体系的升级换代,由原来的自上而下的由少数资源控制者集中控制主导的互联网体系转变为自下而上的由广大用户集体智慧和力量主导的互联网体系。Web2.0 内在的动力来源是将互联网的主导权交还个人从而充分发掘了个人的积极性参与到体系中来,广大个人所贡献的影响和智慧和个人联系形成的社群的影响就替代了原来少数人所控制和制造的影响,从而极大地激发了个人的创作和贡献的潜能,使得互联网的创造力上升到了新的量级。只

收稿日期:2006-09-25

基金项目:国家自然科学基金资助项目(60273043);安徽省自然科学基金资助项目(050420204);安徽省教育厅自然科学基金项目(2006KJ098B, KJ2007B153)

作者简介:李学俊(1976-),男,安徽六安人,讲师,博士研究生,研究方向为机器学习、Web 挖掘;李龙澍,博导,教授,研究方向为智能软件、知识工程、软件体系结构。

有在 Web 2.0 时代,网民才真正成为信息资源的控制者,拥有平等地位;才能想写就写(Blog)、想唱就唱(Podcasting)、想看就看(RSS)、想找就找(SNS)、想编就编(Wiki)、想搞怪就搞怪(Numa Numa Dance、MOP、涂鸦),才彻底摆脱了网络巨头们的阴影,成为网络真正的主人,从网“虫”摇身而成网“人”^[2,3]。Web2.0 的重要特征如下:

- (1)互联网是平等交流的平台而不是利用互联网来统治和控制信息资源;
- (2)充分重视并利用集体力量和智慧;
- (3)将数据变成“Intel Inside”;
- (4)分享和参与的架构驱动的网络效应;
- (5)通过带动分散的、独立的开发者把各个系统和网站组合形成大汇集的改革;
- (6)通过内容和服务的联合使轻量的业务模型可行,分享经济的模式;
- (7)注重用户体验的持续的服务;
- (8)服务和应用无处不在,非单机版和单一平台版本。

2 Wiki 简介

Wiki 一词来源于夏威夷语的“wee kee wee kee”,原本是“快点快点”的意思,被译为“维基”或“维客”。Wiki 指一种超文本系统实现技术,用以实现超文本系统。这种超文本系统支持面向社群的协作式写作,同时也包括一组支持这种写作的辅助工具。可以在 Internet 的基础上对 Wiki 页面进行浏览、创建、更改,而且创建、更改、发布的代价远比 HTML 文本小;同时 Wiki 系统还支持面向社群的协作式写作,为协作式写作提供必要帮助;最后,Wiki 的协作者自然构成了一个社群,Wiki 系统为这个社群提供简单的交流工具。与其它超文本系统,如 Blog, BBS 相比, Wiki 有使用方便及开放的特点,所以 Wiki 系统可以帮助人们在一个社群内共享某领域的知识,构建知识社区共享平台。

随着 Wiki 社区的应用发展,已经出现了很多支持该社区的 Wiki 系统。按照系统运行环境,主要分为四类:基于 PHP 环境的 MediaWiki, CooCooWakka 等;基于 JSP 环境的 JspWiki 等;基于 CGI 环境的 TWiki, UseModWiki;基于 ASP 环境的 JsWiki, OperatorWiki, OpenWiki 等。

目前 Wiki 系统处于初始发展阶段,各种系统具有不同的功能和特点,并没有统一的标准。在众多 Wiki 系统中,MediaWiki 是全球用户最多的、最大的、开源的 Wiki 系统,全世界最大的 Wiki 项目维基百科是 MediaWiki 的成功范例。因此文中基于现有的 Medi-

aWiki1.5rc4 系统,分析研究 Wiki 语言的实现机制和 Wiki 系统的实现机制及其功能。

3 Wiki 语言的实现机制

Wiki 系统的页面编辑功能不同于专门的网页编辑软件,如 FrontPage, Dreamweaver 等,而是采用特有的页面编辑语言 Wiki 语言,给网民提供一个编写平台。该语言类似于 HTML 语言中的标记语言,用于描述文本的样式、字体与页面设置等,但不像 HTML 中,需要标记及其属性的组合,才能完成页面编辑, Wiki 语言编辑更方便、更高效^[4]。Wiki 语言的实现机制,如图 1 所示。

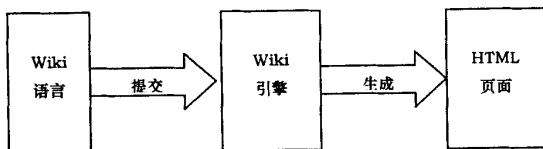


图 1 Wiki 语言的实现机制图

下面以与 HTML 语言作比较的方式,介绍几种有代表性的 Wiki 语言格式。

(1)标题。

= 文档标题 =, 表示一级标题; == 文档标题 ==, 表示二级标题; === 文档标题 ===, 表示三级标题, 等等, 依次类推。

而 HTML, <H1>, <H2>, <H3>, ..., <H6> 双标记仅表示是六种不同等级的标题标记, 从大到小。例如 <H1> 文档标题 </H1>, 表示一级标题。

(2)超级链接。

语法规则: [链接名称 | 链接名称链接目标(文档)], 表示超级链接。例如要表示超级链接: DoNews Wiki 国内最大的 IT 知识库, 可以表示为: [DoNews Wiki 国内最大的 IT 知识库 | http://wiki.donews.com/]。

而 HTML, 表示为: DoNews Wiki 国内最大的 IT 知识库 。

(3)插入图片。

语法规则: [[Image:图像文件 | 参数 1 | 参数 2 | 说明文字]], 表示插入图片标记。例如, 在页面中显示一副图像, 将宽度设为 100px, 可以表示为: [[Image: WikiLogo.jpg | 100px]]。

而 HTML, 表示为:

(4)表格。

语法规则: || 表格参数 ||, 表示 Table 标记; 语法

格式:|“单元格参数”|单元格内容,表示 TD 标记;TH 标记与 TD 类似;语法格式:| - “参数”,表示 TR 标记。例如页面显示表格:

火狐	Firefox
雷鸟	Thunderbird

可以表示为:| border = 1 | 火狐 | Firefox | - | 雷鸟 | Thunderbird | |。

而 HTML,表示为:<table border = 1><tr><td>火狐</td><td>Firefox</td></tr><tr><td>雷鸟</td><td>Thunderbird</td></tr></table>。

4 Wiki 系统的实现机制及其功能

基于 Wiki 系统平台,网民使用 Wiki 语言编辑页面,Wiki 引擎把 Wiki 页面转换成 HTML 页面,返回到客户端,同时自动记录 Wiki 页面的编辑版本,构建成 Wiki 知识库。Wiki 系统的实现机制,如图 2 所示。

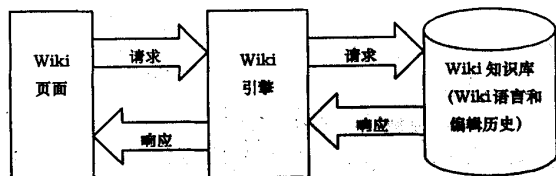


图 2 Wiki 系统的实现机制图

为了实现在 Wiki 知识库中存储 Wiki 语言形式的知识以及编辑历史信息,需要设计相应的数据库结构,主要涉及到三个表:WikiPage(Wiki 页面)、Revision(编辑历史信息)、WikiText(Wiki 语言形式的知识)。每个表的简要描述如下:

(1) WikiPage: PageID(Wiki 页面编号)、PageTitle(Wiki 页面标题)。

(2) Revision: RevID(历史版本编号)、PageID(Wiki 页面编号)、TextID(Wiki 知识编号)、UserID(用户编号)、TimeStamp(编辑时间)。

(3) WikiText: TextID(Wiki 知识编号)、Text(Wiki 知识)。

基于上述数据库结构,Wiki 系统可以实现:Wiki 页面编辑、版本比较、知识分类、系统管理等功能^[5,6]。下面将详细研究 Wiki 系统的主要功能。

4.1 Wiki 页面编辑

通过在线帮助,网民可以使用 Wiki 语言编辑页面。为了网民方便、高效地编写知识,系统也提供了类似于 BBS 中的图形化编辑工具栏。编辑结束后,网民可以先预览编辑效果,确认结果后,系统保存 Wiki 语言形式的知识。系统把编辑历史信息保存到 Revision

表,Wiki 知识保存到 WikiText 表。

类似于 BBS、博客,Wiki 页面可以嵌入图片,但需要先把图片文件以 FTP 或 HTTP 方式上传到服务器指定目录,同时把图片信息保存到数据库中,然后使用 Wiki 语言插入图片。一般情况下,一次只能上传一张图片文件,为了解决一次上传指定目录中的多张图片文件,在本系统的基础上,新增了一次上传多张图片文件的功能。该功能的主要实现过程如下:

- ①获取用户设定目录中上传的文件数目;
- ②以循环方式处理所有文件,对每个文件判断上传是否成功,出错则跳过,继续循环;
- ③显示上传结果。

4.2 版本比较

网民编辑 Wiki 页面,Wiki 知识保存到 WikiText 表,同时版本编辑信息,如编辑者、编辑时间、所编辑页面等,保存到 Revision 表。这样网民可以查看某一 Wiki 页面的所有历史版本,也可以比较所有版本之间的差别,可以恢复到任一历史版本。

系统比较当前 Wiki 知识与历史版本之间的差别,采用不同的颜色表示不同的编辑情况。默认情况下黑色字符表示相同的行,红色加亮表示不同的行,黄色加亮表示删除的行,绿色加亮表示添加的行。该功能基于经典的最长公共子序列(LCS, Longest Common Subsequences)算法实现的。

Wiki 知识库存储所有网民编辑的 Wiki 知识,这样网民可以恢复到任一历史版本。然而管理员也可以把某些不需要修改的 Wiki 页面锁定,以禁止编辑。

4.3 知识分类

系统支持知识分类,并根据知识分类标识在不同知识之间自动产生关联。这样网民在浏览网络知识库时,就可通过页面链接看到相关的知识。可以把机器学习中的分类学习技术如神经网络、归纳学习、关联规则、粗糙集等,应用于 Wiki 知识的自动分类,使 Wiki 网络知识库中的知识有规律、有组织地联系在一起。

4.4 系统管理

Wiki 知识平台是共享、自由平等的,网民可以随意更改任意页面。为了防止某些网民蓄意乱改 Wiki 知识,系统也采用了类似于 BBS 的系统管理功能,如帐户管理、IP 地址管理等。这样当管理员发现到网民的蓄意乱改后,就可以恢复到正确的版本,并禁用帐户或封 IP 地址。但是这种管理方法是人工操作的,具有滞后性。因此可以考虑数据挖掘技术应用于网络知识平台和网民行为的检测分析,及时发现问题,及时更正,这样的系统管理就是自动化的、实时性的。系统也提供远程管理功能,方便管理员远程管理 Wiki 系统。

5 总 结

笔者基于现有的运行于 PHP + MySQL 环境的 MediaWiki . 5rc4 系统,研究了 Web2.0 的典型应用 Wiki 系统,包括 Wiki 语言的实现机制和 Wiki 系统的实现机制及其功能,并设计实现了改进的 Wiki 系统

——Oracle 网络知识库,如图 3 所示。

Web2.0 革命序幕刚刚开始,已经产生了一些基于 Web2.0 的应用。但是 Web2.0 标准仍有待研究,Web2.0 的应用仍有待发展与完善。文中仅研究与应用了 Web2.0 的应用——Wiki 系统,其它的应用如博客、RSS,仍有待研究与应用。随着基于 Wiki 系统的网络知识库在各个知识领域的应用,将会促进 Wiki 系统和 Web2.0 的应用与发展。

参考文献:

- [1] 金 磊. 中国首届 Web2.0 年会专题[EB/OL]. 2006 - 04 - 03. <http://tech.sina.com.cn/focus/webforum-china/>.
- [2] Chen A. Web2.0 周刊[EB/OL]. 2005 - 08 - 04. <http://www.web2week.com/>.
- [3] 奥莱理 T,玄伟剑. 什么是 web2.0[J]. 互联网周刊,2005 (40):38 - 40.
- [4] 陈欣惠. 天下维客[EB/OL]. 2005 - 06 - 05. <http://www.allwiki.com/>.
- [5] Wilkinson R. MediaWiki[EB/OL]. 2006 - 05 - 09. <http://www.mediawiki.org/>.
- [6] 孟 明. MediaWiki 引擎源码下载[EB/OL]. 2006 - 08 - 10. <http://www.cncode.com/downinfo/121.html>.



图 3 基于 Wiki 系统的 Oracle 网络知识库

2007 年全国高性能计算学术年会征文通知

(National Annual Conference on High Performance Computing, HPC China 2007)

2007 年 10 月 26~28 日深圳·明华国际会议中心

<http://www.siat.ac.cn/hpcchina2007>

由中国计算机学会高性能计算专业委员会主办,中国科学院深圳先进技术研究院承办的“2007 年全国高性能计算学术年会”将于 2007 年 10 月 26 日至 28 日在深圳召开。全国高性能计算学术年会是中国高性能计算领域一年一度的盛会,为相关领域的学者提供交流合作、发布最前沿科研成果的平台,并将有力促进中国高性能计算的发展。会议的学术内容及展览涵盖高性能计算应用特征分析、数据模拟与仿真、高性能计算机系统、网格应用、并行算法等多个领域。本届会议的主题是展示和推广高性能计算技术在各学科和领域的应用。欢迎从事高性能计算及相关研究的专家、学者踊跃投稿。

投稿须知:

1、论文应是未发表的研究成果,语言要求中文,格式请参照《计算机研究与发展》(<http://crad.ict.ac.cn/fkz-gya.htm>)。

2、会议录用论文将在《计算机研究与发展》(正刊/增刊)或会议论文集上发表。

3、论文通过电子邮件提交至: hpcchina2007@siat.ac.cn

4、论文提交截止时间:2007 年 08 月 26 日

联系方式:

中国科学院深圳先进技术研究院 朱定局

电话:0755 - 26803533 13266867589

E-mail: dj.zhu@siat.ac.cn; zhudingju@ncic.ac.cn