

# 基于 Web 2.0 的信息服务系统

杨为民<sup>1</sup>, 李龙澍<sup>2</sup>

(1. 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039;

2. 安徽大学 计算机科学与技术学院, 安徽 合肥 230039)

**摘 要:** Web 2.0 是相对 Web 1.0 的新的一类互联网应用的统称。基于 Web 2.0 模式, 用户可以主动创造互联网信息, 实现互联网的交互、互动、参与的目标。构建一个面向特定的主题的、基于 Web 2.0 的构件化信息服务平台。它能更好地服务于用户个体的微内容的收集、创建、发布、管理、分享、合作、维护等, 对微内容进行表现、聚合、迁移, 对社会性关系进行维护。

**关键词:** 信息服务系统; Web 2.0; 微内容; 智能信息处理

**中图分类号:** TP391.1

**文献标识码:** A

**文章编号:** 1673-629X(2007)09-0181-04

## An Information Service System Based on Web 2.0

YANG Wei-min<sup>1</sup>, LI Long-shu<sup>2</sup>

(1. Ministry of Education Key Lab. of Intelligent Computing & Signal Processing,

Anhui University, Hefei 230039, China;

2. School of Computer Science and Engineering, Anhui University, Hefei 230039, China)

**Abstract:** Web 2.0 is a new Internet application, which is relative to Web 1.0. Users can initiatively release Web information, that means Internet interaction, participation comes true. Building a component information service platform based on Web 2.0, which faced on a specific topic, can collect, create, release, manage, participate, cooperate, maintain individual microcontent. It also can represent, aggregate, transfer the microcontent and maintain the social relation.

**Key words:** information service system; Web 2.0; microcontent; intelligent information processing

## 0 引言

Web 2.0 这一概念, 由 O'Reilly 媒体公司总裁兼 CEO 提姆·奥莱理提出。Web 2.0 目前没有一个统一的定义。这里只给出 Web 2.0 的特征描述。它是以 Flickr, Craigslist, LinkedIn, Tribes, Ryze, Friendster, Del.icio.us, 43Things.com 等网站为代表, 以站点摘要 RSS、博客/网志 Blog、百科全书 Wiki、网页书签 TAG、社会网络 SNS、对等联网 P2P、即时通讯 IM 等应用为核心, 依据六度分隔、XML、Ajax 等新理论和技术实现的互联网新一代模式<sup>[1]</sup>。目前, RSS 广泛用于 Blog, Wiki 和网上新闻频道, 世界多数知名新闻网站都提

供 RSS 订阅支持, Blog 托管服务商 (BSP) 也有很多, Wiki 系统、TAG、SNS 已在许多站点可以使用, P2P 和 IM 是目前我国上网用户使用率极高的软件。

笔者提出基于 Web 2.0 的构件化信息服务平台。

## 1 Web 2.0

Web 2.0 是相对 Web 1.0 (2003 年以前的互联网模式) 的新的一类互联网应用的统称, 是一次从核心内容到外部应用的革命<sup>[2]</sup>。从 Web 1.0 到 Web 2.0 的进化实现了在模式上由单纯的“读”向“写”、“共同建设”发展; 对于用户, 则实现了由被动地接收互联网信息向主动创造互联网信息迈进; 从基本构成单元上, 是由“网页”向“发表/记录的信息”发展; 从工具上, 是由互联网浏览器向各类浏览器、RSS 阅读器等内容发展; 从运行机制上, 由“Client Server”向“Web Services”转变; 作者由程序员等专业人士向全部普通用户发展; 应用上由初级的片面应用向全面大量应用发展。

Web 2.0 的基本要素是微内容、用户个体、社群分

收稿日期: 2006-11-11

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); 安徽省高校拔尖人才基金资助项目; 安徽省教育厅自然科学基金项目 (2006KJ098B)

作者简介: 杨为民 (1968-), 男, 安徽合肥人, 博士研究生, 研究方向为智能信息处理、信息检索和智能软件; 李龙澍, 博士生导师, 教授, 研究方向为知识工程、智能软件和软件工程。

享和 API 编程接口。

(1)微内容(microcontent): 用户所生成的任何数据,比如评论、图片、喜爱的音乐列表等。这些微内容的数量、重要性,以及用户对它的依赖,不亚于正统文章、论文、书籍。对微内容的重新发现和利用,充分体现了互联网的平等、自由。Web 2.0 的关键

技术就是帮助用户管理、维护、存储、分享、转移微内容。

(2)用户个体:对于 Web 1.0 的典型产品/服务来说,用户没有具体的特征个性,它只是一个模糊的群体的代名词。对于 Web 2.0 的产品和服务来说,用户是个实实在在的人。Web 2.0 所服务的对象是具体的人,而且人的具体性会随着服务不断地充实起来。Web 2.0 设计的起点是为这个具体的个体服务的。

(3)社群分享:Web 社群的主体内容是构成社群的用户。主体内容又能够经自由组合从而衍生新的社群。也就是说,用户能够借助内容这一媒介,自由地创建起一个个的社群,发生种种社会性的(网络)行为。

(4)应用接口:在 Web 2.0 阶段,Web 是一个平台,借助 Web 服务提供的编程接口,网站成了软件构件。目前,象 del.icio.us, flickr, a9, amazon, yahoo, google, msn 等网站提供的编程接口,可以衍生出各种应用。

## 2 基于 Web 2.0 的信息服务平台

信息服务平台(见图 1)采用 ADL 和 UML 相结合的体系描述方法,嵌入了构件化的 Web 2.0 典型应用系统(如 Blog 网络博客、Wiki 百科全书、IM 即时通讯等);在统一的信息描述结构下,该平台实现了构件之间的信息共享与功能集成;针对海量微内容的信息冗余性,采用模糊 Rough 集进行信息过滤;利用商空间理论设计信息粒度处理算法和交叉覆盖聚类算法,自动生成面向不同用户个体需求的、不同粒度的 RSS 微内容信息摘要;采用“推送”的方式,实现用户的信息定制,并向用户及时推荐适用信息。

### 2.1 Web 2.0 的典型应用

#### (1)Wiki。

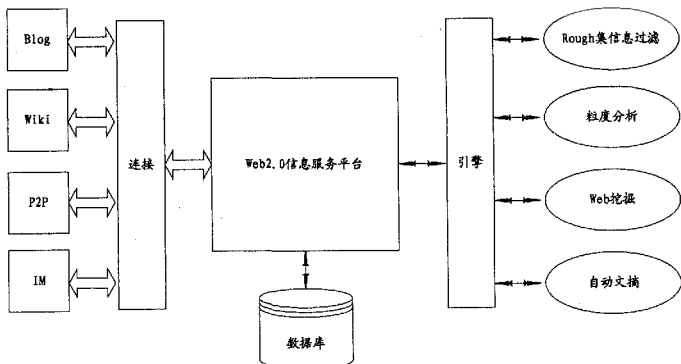


图 1 基于 Web 2.0 的信息服务平台

Wiki(夏威夷语,意为 wee kee wee kee)是一种超文本系统。它支持面向社群的协作式写作,包括一组支持这种写作的辅助工具。用户可以在 Web 的基础上对 Wiki 文本进行浏览、创建、更改。这种创建、更改、发布的代价远比 HTML 文本为小。Wiki 系统支持面向社群的协作式写作,为协作式写作提供必要帮助。这样,Wiki 的写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。与其它超文本系统相比,Wiki 使用方便、开放,可以帮助用户在一个社群内共享某领域的知识。

#### (2)Blog。

Blog(Web log),称为网志、部落或者博客,是一种网络共享空间,也是一种以日记的形式在网络上发表自己的个人内容的形式。Blog 是由简短且经常更新的 Post 所构成。文章都按照年份和日期排列。Blog 的内容极为丰富,从对其他网站的超级链接和评论,有关公司、个人、构想的新闻到日记、照片、诗歌、散文,甚至科幻小说的发表或张贴都有,涉及各行各业。大多数的 Blog 是个人心中所想之事情的发表,倾向于个人情感的体现活动。也有一些 Blog 是一群基于某个特定主题或共同利益领域的集体创作。Blog 注重网络传达的实时讯息。可以说 Blog 是继 Email, BBS, ICQ 之后出现的第四种网络交流方式。这是一种新的生活方式、工作方式、学习方式和交流方式,是“互联网的第四块里程碑”<sup>[2]</sup>。

#### (3)P2P。

P2P(peer-to-peer)是“伙伴对伙伴”的意思,或称为对等联网。它对加强网络上人的交流、文件交换、分布计算等方面作用巨大。P2P 直接将人们联系起来,让人们通过互联网直接交互。P2P 使得网络上的沟通变得更容易、更直接。通过共享和交互,真正地消

除中间商。P2P 就是人可以直接连接到其他用户的计算机。这样,交换文件就不需通过服务器浏览与下载。P2P 的另一个重要特点是改变互联网现在的以大网站为中心的状态、重返“非中心化”,并把网络的权力交还给用户。

#### (4) IM。

IM(Instant Messenger,即时通讯)软件可以说是目前我国上网用户使用率最高的软件。聊天一直是网民们上网的主要活动之一。网上聊天的主要工具已经从初期的聊天室、论坛变为以 MSN、QQ 为代表的即时通讯软件。大部分人只要上网就会开着自己的 MSN 或 QQ。作为使用频率最高的网络软件,即时聊天已经突破了技术工具的极限,被认为是现代交流方式的象征。它也构建了一种新的社会关系。即时聊天是迄今为止对人类社会生活改变最为深刻的一种网络新形态:没有极限的沟通带来没有极限的生活。

将众多 Web 2.0 的应用系统构件化,通过统一的接口,实现构件间功能的无缝集成,实现所有的构件间的信息共享。

### 2.2 构件与连接

构件是软件系统中具有某种功能可重用的结构块单元,它包括计算、数据存储、接口、约束等部分<sup>[3]</sup>。构件可以分为:数据服务构件、功能服务构件、界面构件、控制构件、连接构件等等。不包含其他构件的构件称为原子构件,由其他构件连接构成的构件称为复合构件。Client, server, filter, interpreter, scheduler, database 等等都是典型的构件<sup>[4]</sup>。构件是进行计算或数据存储的单元,包括以下几部分:接口(interface)、类型(types)、语义(semantics)、约束(constraints)、进化(evolution)、非功能特性(non-functional properties)等。构件是一个数据单元或一个计算部件,它由构件接口和构件实现模块组成<sup>[5]</sup>。构件接口是构件与外部端口的集合。每一个端口包括构件的标识、端口提供给环境或其它构件功能集合、端口运行所需环境或其它构件的功能集合、端口的私有属性集合、端口的行为语义描述、端口所产生消息的集合、对构件端口行为约束以及端口的非功能说明,包括构件的安全性、可靠性说明等。

连接是构件之间的接口模块,它包含了信息交换的规则<sup>[5]</sup>。就是说连接包括实现机制和信息交换协议两部分。常见的连接如:栈、缓冲区、解释器、管道(pipes)、过程调用、共享、线程、并行等等,更为复杂的连接如:client-server 通信协议、数据库和应用之间的 SQL 连接、ActiveX、ODBC 等等。复杂的连接需要专门的结构来完成,连接也可以看作是一种构件。

连接一般包括如下几部分:接口(interface)、类型(types)、语义(semantics)、约束(constraints)、进化(evolution)、非功能特性(non-functional properties)等。

### 3 基于商空间的信息处理

信息服务平台对各构件所获得的微内容信息进行统一的信息描述;针对海量微内容的信息冗余性,运用粗糙集(Rough Set)理论进行信息过滤;利用商空间(Quotient Space)理论的信息处理算法,对信息进行聚类(Clustering)、分类(Classification),自动生成面向不同用户个体需求的、不同粒度(Granularity)的 RSS 微内容信息摘要;采用“推送”的方式,实现用户的信息定制。

#### 3.1 统一的信息表示

XML(Extensible Markup Language,可扩展标记语言)是一种可以用来创建自己的标记的标记语言<sup>[6]</sup>。它由万维网协会(W3C)创建,用来克服 HTML(Hypertext Markup Language,超文本标记语言)的局限。XML是基于 SGML(Standard Generalized Markup Language,标准通用标记语言)的用来定义其他语言的元语言。它是表示 Web 上结构化信息的一种标准文本格式,既无标签集也无语法。XML 有许多优点:

(1)可扩展性。XML 允许使用者创建和使用他们自己的标记而不是 HTML 的有限词汇表。企业可以用 XML 为电子商务和供应链集成等应用定义自己的标记语言。特定行业也可以定义其行业领域的特殊标记语言,作为其领域信息共享与数据交换的基础。

(2)灵活性。因为 HTML 是格式、超文本和图形用户界面语义的混合,要进一步发展这些混合在一起的功能很困难。而 XML 提供了一种结构化的数据表示方式,使得用户界面分离于结构化数据。所以,Web 用户所追求的许多先进功能在 XML 环境下容易实现。

(3)自描述性。XML 文档通常包含一个文档类型声明,它是自描述的。不仅人可以读懂 XML 文档,计算机也能直接进行处理。XML 表示数据的方式真正做到了独立于应用系统,而且数据能够重用。因此,XML 文档被看作是文档的数据库化和数据的文档化。

(4)简明性。XML 大约只有 SGML 复杂性的 20%,却具有了 SGML 功能的 80%。它比完整的 SGML 简单得多。XML 易学、易用且易实现。

此外,XML 吸收了人们多年来在 Web 上使用 HTML 的经验。XML 支持世界上几乎所有的主要语言,并且不同语言的文本可以在同一文档中混合使用,应用 XML 的软件能处理这些语言的任何组合。所有

这一切将使 XML 成为数据表示的一个开放标准,这种数据表示独立于机器平台、供应商以及编程语言。

### 3.2 基于商空间的

#### 信息处理

商空间理论是

张铃和张钺提出的

粒度计算理论<sup>[7]</sup>。该理论认为概念可以用子集来表示,不同粒度的概念体现为不同粒度的子集,一族概念就构成空间的一个划分——商空间(知识基)。不同的概念簇就构成不同的商空间。因此,粒度计算就是研究在给定知识基上的各种子集合之间的关系和转换,以及对同一问题,取不同的适当的粒度,从对不同的粒度的研究中,综合获取对原问题的了解。

商空间的模型可用一个三元组来表示,即  $(X, F, T)$ , 其中  $X$  是论域,  $F$  是性质集,  $T$  是  $X$  上的拓扑结构。取粗粒度即是给定一个等价关系  $R$  (或说一个划分), 从而得到一个对应于  $R$  的商集记为  $[X]$ 。它对应于的三元组为  $([X], [F], [T])$ , 被称为对应于  $R$  的商空间。商空间理论就是研究各商空间之间的关系, 各商空间的合成、综合、分解和在商空间中的推理。波兰学者 Pawlak 在 20 世纪 80 年代提出了粗糙集(Rough Set)理论<sup>[8]</sup>。他提出一个假设: 人的智能(知识)就是一种分类的能力, 这个假设可能不是很完备, 但却非常精练。在此基础上提出, 概念可以用论域中的子集来表示, 于是在论域中给定一组子集族, 或说给定一个划分(所谓划分, 是指将  $X$  分成两两不相交的子集之并)。

给定  $X$  上的一个划分, 等价于在  $X$  上给定一个等价关系  $R$ 。  $R$  即为在论域上给定了一个知识基  $(X, R)$ 。对一般的概念  $x$  ( $X$  中的一个子集), 可以用知识基中的知识来表示, 即用知识基中的集合的并来表示。对那些无法用  $(X, R)$  中的集合的并来表示的集合, 可以借用拓扑中的内核和闭包的概念, 引入  $R_-$  下近似  $R_-(x)$  (相当于  $x$  的内核) 和  $R_-$  上近似  $R_+(x)$  (相当于  $x$  的闭包)。当  $R_-(x) \neq R_+(x)$  时, 就称  $x$  为粗糙集。

目前, 商空间理论和粗糙集理论已被广泛应用于各个领域, 特别是数据挖掘领域, 并获得成功。

### 3.3 信息处理流程

通过对构件所获得的微内容信息, 通过引擎以 XML 的格式进行统一的表示; 运用粗糙集理论对统一表示的海量微内容信息进行信息过滤, 消除信息冗余; 运用商空间的粒度计算理论, 对微内容信息进行聚类分类; 结合用户个性化信息需求特征, 自动生成面向不同用户个体需求的、不同粒度的 RSS 微内容信息摘

要; 采用“推送”的方式, 实现用户的信息定制, 并向用户及时推荐适用信息。信息处理流程见图 2。

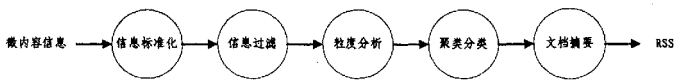


图 2 信息处理流程

## 4 结束语

基于 Web 2.0 的构件化信息服务平台, 在面向特定的主题时, 如古代建筑, 能服务于用户个体(专业人员、一般用户)的微内容的收集、创建、发布、管理、分享、合作、维护等, 对微内容进行表现、聚合、迁移, 对社会性关系进行维护。

具体地说, 信息服务平台可以根据古代建筑语义表达的法则进行主题分类、信息约简, 运用内容挖掘、结构挖掘和日志挖掘, 实现 RSS 整合型内容的推送、信息搜索、个性化以及音乐、短视频的高效快速访问; 可以从海量的微内容信息中抽取古代建筑的知识, 按照古代建筑的年代、地点、型制、结构体系、构造特征等属性进行知识重构, 实现知识发现; 提供古代建筑的信息库, 包括知识库、数据库和模型库。

进一步的工作是对服务平台的微内容信息进行深层的数据挖掘, 以便产生新的信息、知识, 引导用户的兴趣趋向。

### 参考文献:

- [1] Tim O'Reilly. What Is Web 2.0 [EB/OL]. 2005-09. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-2.0.html>.
- [2] Don. WEB 2.0 概念诠释 [EB/OL]. 2005-12. <http://it.sohu.com/s2005/web2info.shtml>.
- [3] Medvidovic N, Taylor R N. A Classification and Comparison Framework for Software Architecture Description Languages [J]. IEEE Transactions on Software Engineering, 2000, 26 (1): 70-93.
- [4] Shaw M, Garland D. Software Architecture: Perspectives on an Emerging Discipline [M]. Beijing: Tsinghua University Press, 1998.
- [5] 赵会群, 王国仁, 高 远. 软件体系结构抽象模型 [J]. 计算机学报, 2002, 25(7): 730-736.
- [6] Jake Sturm. Developing XML Solutions [M]. Washington: Microsoft Press, 2000.
- [7] 张 铃, 张 钺. 基于商空间模型的粒度计算 [J]. 软件学报, 2003, 14 (4): 770-776.
- [8] Pawlak Z. Rough Sets Theoretical Aspects of Reasoning about Data [M]. London: Kluwer Academic Publishers, 1991.