

主动信息服务系统的设计与实现

王 肃¹, 杜军平²

(1. 北京工商大学 计算机学院, 北京 100037; 2. 北京邮电大学 计算机学院, 北京 100876)

摘 要:针对旅游业的现状,开发了主动信息服务系统,给出了智能信息推拉技术的特点和种类,建立了推送模型和拉模型.提出了基于智能信息推拉技术的主动信息服务系统的设计方案,给出了系统开发的流程图.采用用户兴趣模型和可信度模型相结合的分析方法采集网络信息数据,利用 C#.net 编程语言和 Ajax 技术开发了主动信息服务系统.该系统达到了平衡用户需求和信息可信的目的,具有一定的实用价值.

关键词:智能信息推拉; 旅游; 主动信息服务

中图分类号: TP311 **文献标识码:** A **文章编号:** 1673-4785(2007)06-0072-05

Design and implementation of an active information service system

WANG Su¹, DU Jun-ping²

(1. School of Computer Science, Beijing Technology and Business University, Beijing 100037, China; 2. School of Computer and Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Facing the situation of tourism industry, the active information service system is developed. This paper discusses the types and characteristics of intelligent information push and pull technology (IIPP) and establishes a push model and a pull model, respectively. A design scheme of active information service system based on intelligent information push and pull technology is proposed. The flowchart of system development is given. The network data are collected by using both the interest model and credibility model. In the process of system development, C#.net programming language and Ajax technology are involved. The system balances the user's interest requirements and data credibility.

Keywords: intelligent information push and pull; tourism; active information service

旅游业已经成为 Internet 上最普遍最活跃的技术服务项目之一,旅游者对信息服务的依赖程度也越来越高.在出行前作旅行决策时,旅行者需要借助各种媒介(旅游报刊、互联网)了解各地的旅游信息,通过媒介提供的文字、图片、视频等各种综合信息来作决策.由于网络信息资源的分散性和针对性,旅游者很难根据特定的信息源作为自己的参考,因此需要主动信息服务系统来提供及时、准确和多样化的信息^[1-2].

主动信息服务系统的特性主要体现在运用模型分析用户需求、搜集相关信息、信息处理和主动推送、减轻用户的手工操作等.因而在实现上,势必要

跟踪、分析用户的兴趣需求、过滤次要信息、反映重要的信息.智能信息推拉是主动服务的技术实现,它按一定的方法和技术改善了系统与用户的交互.在结构上,智能信息推拉(intelligent information push and pull, IIPP)将推(Push)与拉(Pull)2种信息获取方式进行有机的结合,其目的是根据 Pull 的历史趋势,加强 Push 的针对性;同时利用 Pull 历史和 Push 的功能,提高 Pull 的效率^[3-4].本文给出了基于智能信息推拉技术的旅游主动信息服务系统的设计,并开发出旅游网站的主动信息服务系统.

1 智能信息推拉(IIPP)技术

智能信息推拉技术是传统信息“推送”与“拉取”技术的结合,在两者的基础上再融入人工智能、知识

收稿日期:2007-08-07.

基金项目:国家自然科学基金资助项目(60773112).

发现、Internet 及数据库等技术,从而形成智能信息推拉技术.智能信息推拉技术主要有 4 种方式

1) 先推后拉式.先由信源及时推送公共信息,再由用户有针对性地拉取个性化信息.该模式可适用于图书情报部门的信息服务.例如可将电子期刊的目次信息推送给用户,供其浏览,用户一旦发现自己所需信息,可再连接到服务器浏览或下载全文.

2) 先拉后推式.用户先拉取搜索所需信息,可根据用户的兴趣,再有针对性地推送相关的其他信息,如个性化的频道订阅、邮件订阅、电台节目点播等.该模式既能满足用户个性化需求,又能进行主动服务,所以这种方式在 Internet 上应用较多.

3) 推中有拉式.在信息推送过程中,允许用户随时中断、定格在所感兴趣的网页上,并进一步搜索,主动拉取更丰富的信息.

4) 拉中有推式.在用户拉取信息的搜索过程中,根据用户输入的查询信息,信源主动推送相关信息和最新信息.

2 推送模型的建立

推送模型如图 1 所示.

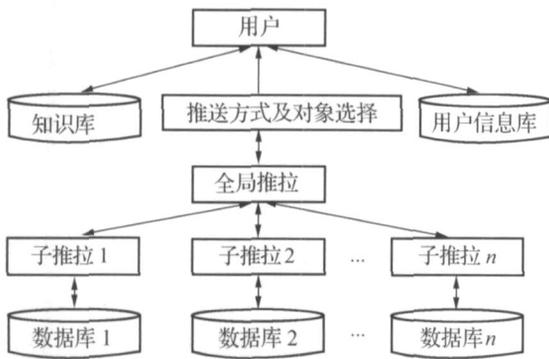


图 1 推送模型

Fig. 1 Push model

该模型采用分布式信息推拉,使用若干个子推拉分解用户的请求,最后通过全局推拉得到最终的结果.在推送方式及对象选择过程中,根据知识库和用户信息库得到用户的基本信息、用户的需求信息、用户的兴趣爱好、有关商品、产品信息等,把分布式信息推拉中得到的结果有选择地发送给相应的用户.推送方式主要分为网页式和电子邮件式.网页式

使用户浏览信息的方式更为灵活,而且可以提供比较丰富的页面信息和类型呈现给用户,用户只需点击或经过简单选择输入即可获得相关信息;还可以根据用户的类型和信息类型,给用户发送电子邮件,除了提供用户需要的信息外,还提供与此相关的其他信息,也可以将产品的广告发送给用户^[5-6].

3 拉取模型的建立

拉取模型如图 2 所示.

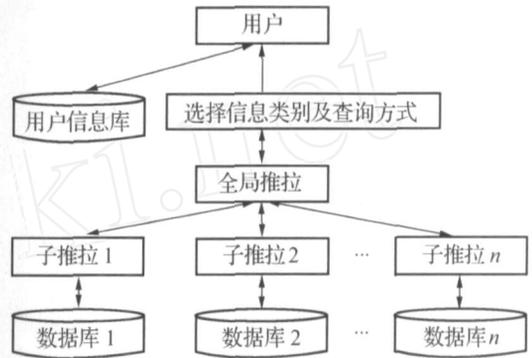


图 2 拉取模型

Fig. 2 Pull model

用户根据自己的需求及兴趣爱好,选择要查询的信息类别,包括酒店预订、车船票预订、询价报价查询、旅游产品查询、商情信息查询等,以及查询的方式,包括按地区、按企业、按最佳推荐等、通过分布式信息推拉从相应的数据库中找到相关信息.在拉取模型中,用户有比较大的自主性,但要求用户对信息类别、查找的关键字等有一定的了解.

4 主动信息服务系统的设计

主动信息服务系统一方面根据用户设定的频道更新和推送数据,另一方面根据注册用户的兴趣选择关键词、搜索和采集用户操作信息,为兴趣模型提供分析数据,同时结合网站的可信度模型,得到网络信息采集的策略,系统将数据处理后,推送到客户端.主动信息服务系统及客户端浏览器结构如图 3 所示.

4.1 主要模块介绍

信息管理模块用户可以管理收藏的网页,设定频道和信息显示的选项.输出管理模块按照用户定义的方式管理和显示数据,按照各种条件排序,显示

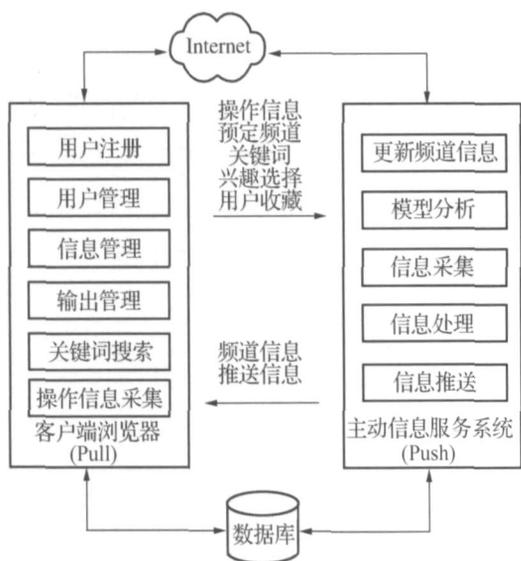


图 3 主动信息服务系统及客户端浏览器结构

Fig. 3 System of active information service and structure of client browser

关键词和其他的页面显示效果。搜索模块使用关键词或者已有的条件组合搜索信息。操作信息采集模块用来采集用户对浏览网页的操作,作为兴趣模型计算的依据。网络信息采集模块根据模型分析得到的权值采集信息,对权值小于阈值的信息做抛弃处理,信息处理模块对网络采集的信息进行过滤和排序。

4.2 分析模型

4.2.1 兴趣模型

首先设定用户操作记录的权值,例如设定下载网页的网站的 IP 次数增长量每超过 10 次,权重自增 0.01,若此时下载某网站的次数为 7,则 $V_1 = 0.01$ 。可以得到用户对网站的兴趣度计算公式:

$$value = \sum_{i=1}^4 V_i^* + oldvalue. \quad (1)$$

式中: V_i 是对应的用户操作从权值表中查询得到的值, i 是此操作的权值,且 $i = 1$ 。

兴趣模型中的所有权值都随时间而减小,一般安排在有效任务执行完成后,并且是系统空闲时进行计算,其计算公式为

$$newvalue = oldvalue \times \left(1 - \frac{MDate}{d} \right). \quad (2)$$

$newvalue$ 是新的权值, $oldvalue$ 是旧的权值, $MDate$ 是上一次访问到现在的天数, d 为调节常数。

4.2.2 可信度模型

Web 搜索引擎使用 Web 的重要性得分来排列

搜索结果,采用页面级别算法来实现。页面级别 (pagerank) 计算系统所遵循的思路是:通过一个互联网的“随机运动”得知哪些站点是最重要的。系统模拟一个随机冲浪者跟进点击某页面上的随机链接,到最深层页面时按“返回”按钮这一过程。pagerank 算法计算一个仅仅基于 Web 链接结构的单一向量,将排名比重转移到了网页以外的因素上,计算链接到某一网页的网站数量,然后按照从 1 ~ 10 分别给予表示重要度的分数。链接到网页的站点越多, pagerank 的分数越高。

主题性页面级别技术 (topic-sensitive pagerank) 首先对查询术语划分类别,然后计算有助于查询主题的 pagerank 得分。主题性页面级别算法预先计算一组 pagerank (页面级别) 向量,每个向量偏向于一个不同的主题。在查询时确定可能的查询主题,根据关于每个主题的 pagerank (页面级别) 向量给网页打分。在系统中利用了 pagerank 作为可信度的参数。

4.2.3 模型的加权

结合兴趣模型和可信度模型,给出了计算页面权值的公式:

$$value = V_1 \times + V_2 \times , \quad (3)$$

式中: V_1 、 V_2 分别是根据兴趣模型和可信度模型得到的权值, $+$ 、 \times 为权值系数,且 $+$ + \times = 1。

4.2.4 反馈及调整

用户对相关网页的浏览其实是对系统的一种隐性反馈,对反馈结果进行分析可以使推送的内容更加符合用户的需求。和 \times 的值用户可以自己设定,也可以由系统计算后逐渐修改权值系数。和 \times 比值的改变反映了系统主动服务的偏向性。

5 系统的开发

采用了 C#.NET 平台,根据图 4 和图 5 的流程进行开发。开发过程中使用了 Ajax 技术,以减少客户端和服务器的通讯量,减轻了服务器的负担,加快了页面的显示速度。实现无刷新更新页面,减少用户实际和心理等待时间。系统原型开发完成后,作为中国旅游目的地营销系统的子系统投入试用。

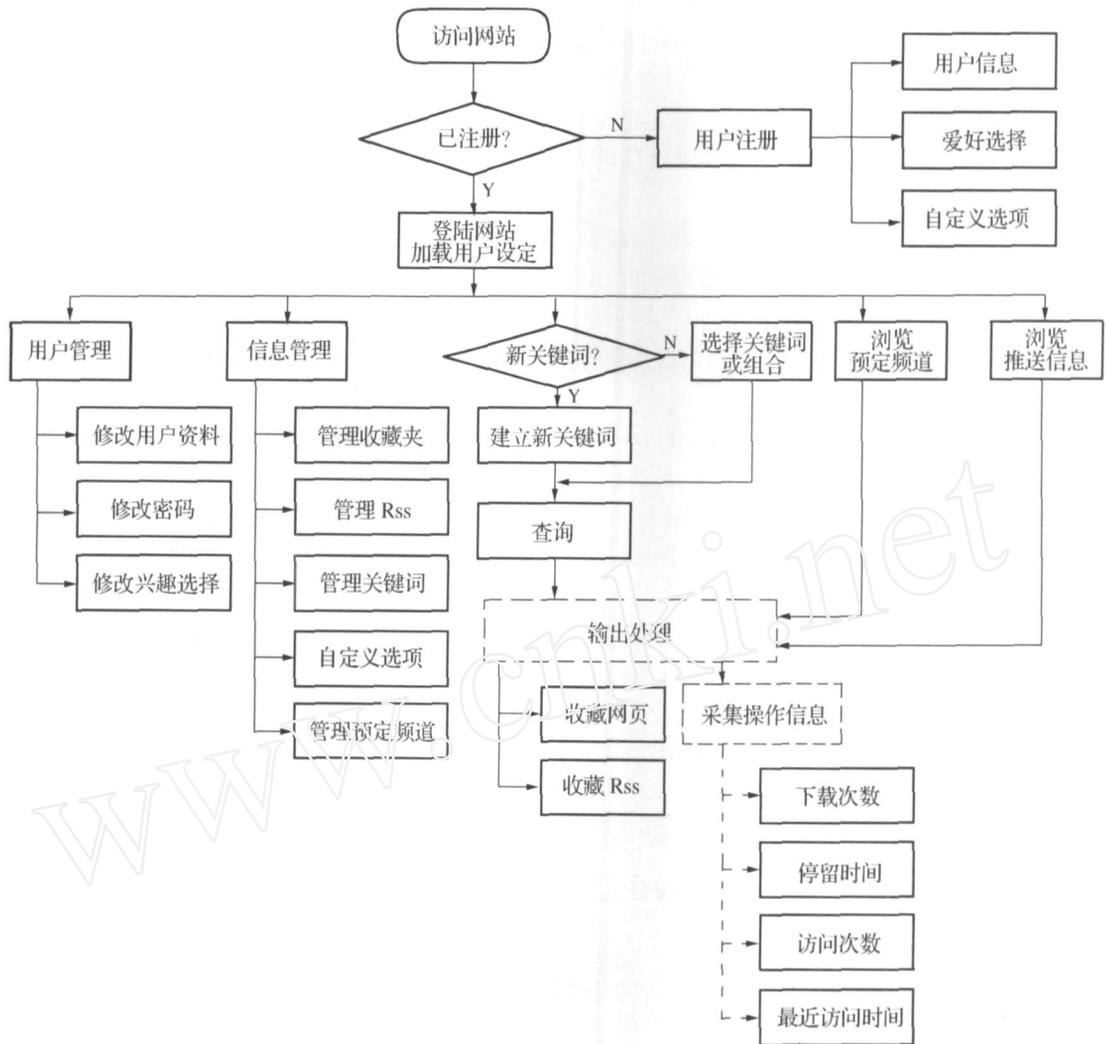


图 4 客户端浏览页面流程图

Fig. 4 Flowchart of client browse web page

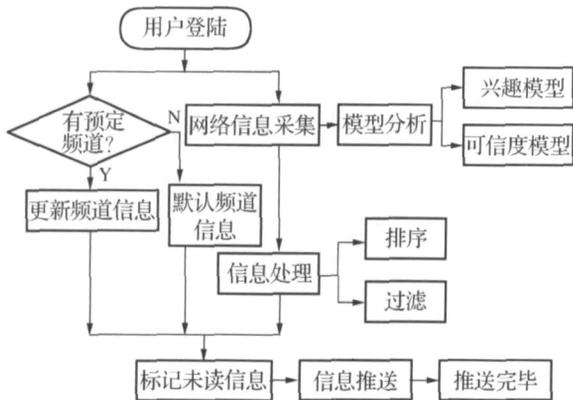


图 5 主动信息服务系统流程图

Fig. 5 Flowchart of active information service

6 结束语

由主动信息服务系统主动提供的信息,不一定完全等同于用户的需求,若要提供恰到好处的服务和信息,主动信息服务系统必须对用户的需求进行

尽可能准确的预测.然而,预测是提供服务者的主观意志,而用户的需求则是客观的存在,预测不可能做到绝对的准确.本系统采用用户兴趣模型和可信度模型,从主观和客观两方面采集网络数据,以达到平衡用户需求和信息可信的目的,具有一定的实用价值.

参考文献:

[1] 邹晓泉. 智能 Agent 及其在电子商务中的应用[J]. 信息与控制, 2003, 32(1): 77 - 81.
 ZHOU Xiaoquan. Intelligent Agent and its applications in e-commerce[J]. Information and Control, 2003, 32(1): 77 - 81.

[2] 焦玉英, 索传军. 基于“推”模式的网络信息服务及其相关技术研究[J]. 情报学报, 2001, 20(2): 193 - 199.
 JIAO Yuying, SUO Chuanjun. Network service mechanism based on the “Push” model and related technologies [J]. Journal of the China Society for Scientific and Technical Information, 2001, 20(2): 193 - 199.

[3] 马红妹, 谭庆平, 陈火旺. 基于 Push 技术的信息获取方式

及其应用[J]. 计算机科学, 1999, 26(3) :55 - 58.

MA Hongmei, TAN Qingping, CHEN huowang. Obtaining information based on Push technology[J]. Computer Science, 1999, 26(3) :55 - 58.

[4]沈 艺. 信息推送技术及其应用[J]. 计算机系统应用, 1999, 10(5) :26 - 27.

SHEN Yi. Information Push technology and its application[J]. Computer Systems & Applications, 1999, 10(5) :26 - 27.

[5]何 军, 周明天. 一种 Push 技术支持的分布信息共享机制[J]. 计算机科学, 1993, 26(7) :71 - 73.

HE Jun, ZHOU Mingtian. A distributed information sharing mechanism supported by another Push technology[J]. Computer Science, 1993, 26(7) :71 - 73.

[6]王忠民, 涂序彦. 推送技术的发展和应[J]. 微计算机信息, 1999, 15(3) :8 - 9.

WANG Zhongmin, TU Xuyan. Development and application of information Push and Pull technology[J]. Control & Automation, 1999, 15(3) :8 - 9.

作者简介:



王 肃,男,1979 年生,硕士研究生,主要研究方向为智能信息处理、计算机网络等.

E-mail:wang2001su@126.com.



杜军平,女,1963 年生,教授,博士生导师,中国人工智能学会常务理事、副秘书长,中国自动化智能自动化专业委员会副秘书长,中国旅游信息标准技术委员会主任委员,主要研究方向为数据挖掘、Agent 理论与技术、智能信息处理、旅游智能信息系统等.近年来完成科研项目 20 余项,发表学术论文 104 篇,出版著作 2 部.

E-mail:junpingdu@bupt.edu.cn.

第 3 届中国可信计算与信息安全学术会议征文通知

Third Conference on Trusted Computing and Information Security

为了加强我国可信计算和信息安全领域学术、技术交流,促进我国可信计算和信息安全领域的学术繁荣、技术进步和产业发展,由解放军密码管理局和中国计算机学会容错专业委员会主办,中国人民解放军信息工程大学电子技术学院承办的第 3 届中国可信计算与信息安全学术会议将拟于 2008 年 10 月 25 ~ 28 日在河南郑州举行.会议将邀请本领域的专家作专题报告,将邀请研究人员、企业代表等针对可信计算与信息安全领域的关键技术和热点问题进行交流和研讨.欢迎各位专家、研究开发及工程技术人员及该领域的企事业人士踊跃投稿参加.

一、征文范围:

会议重点征集可信计算与信息安全理论和技术方面的研究论文,具体包括(但不限于):

(1)可信计算体系结构:可信计算理论,信任理论,可信计算平台体系结构,可信计算软件体系结构,可信网格,容错计算;

(2)可信软件:高可信软件,操作系统安全,数据库安全,软件容错,软件测试;

(3)可信硬件:可信计算平台,可信计算平台模块,信息安全芯片,智能卡,硬件容错,硬件测试,电子设备的物理安全;

(4)网络与通信安全:可信网络,网络安全技术,网络协议安全,网络容侵与容灾,通信安全,无线通信网络安全,计算机病毒技术;

(5)密码学:密码学的理论与技术,新型密码,密码应用技术;

(6)信息隐藏:信息隐藏,数字水印,数字版权管理;

(7)信息安全应用:电子政务安全,电子商务安全,可信计算与信息安全的应,信息安全管理.

二、征文要求:

论文必须为未公开发表且未向学术刊物和其他学术会议投稿的最新研究成果,文稿使用中文或英文书写,字数一般不超过 6000 字.录用的英文稿件将在《武汉大学学报(英文版)》(EI 刊源)上发表,录用的中文稿件在核心期刊《武汉大学学报》(正刊)发表.

三、重要日期:

征文截止日期:2008 年 4 月 30 日;

录用通知:2008 年 6 月 1 日;

返回修改稿:2008 年 6 月 20 日;

定稿时间:2008 年 7 月 1 日.

四、会议投稿方式

本次会议投稿一律通过会议网站 www.tc2008.org 的投稿系统进行.

五、会议通讯地址

(450004)河南郑州商城东路 12 号中国人民解放军信息工程大学电子技术学院信息安全研究所

联系电话:0371-63538081,0371-66094401,0431-38081(军).

联系人:李立新,周雁舟.

E-mail:tc2008_zz@163.com.