

基于 Agent-Aglet 的 Web 信息检索方法

徐朝辉^{1,2}, 秦 杰^{1,2}, 甄 彤^{1,2}

(1 河南工业大学 粮食信息处理与控制教育部重点实验室, 河南 郑州, 450001;

2 河南工业大学 信息科学与工程学院, 河南 郑州, 450001)

摘 要: 本文以 Agent-Aglet 平台为基础, 运用 Java 语言设计了 SearchAglet 并成功应用于 Web 的网络信息检索中, 智能化地将信息检索技术与知识学习技术有机结合, 提出了新的 Web 信息检索方法, 并验证了其可行性和有效性。

关键词: 移动 Agent; Web 信息检索

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)01-0168-04

An Approach of Web Information Retrieval Based on Agent-Aglet

XU Zhao-hui^{1,2}, QIN Jie^{1,2}, ZHEN Tong^{1,2}

(1 Key Laboratory of Provision Information Manage & Control of MOE,

Henan University of Technology, Zhengzhou 450001, China;

2 School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: In this paper, we used Java language to have programmed the “SearchAglet” and successfully applied to web information retrieval. “SearchAglet” intelligently combined the technology of information retrieval with the learning technology, proposed a new approach of Web Information retrieval and verified its feasibility and effectiveness.

Key words: moving Agent; web information retrieval

1 引言

Web 信息资源的大容量、异构性、分布性、动态性的特征, 使得从泛如大海的信息空间中快速而又准确地检索到焦点信息, 已经成为人们期待的目标。Agent 技术的逐步成熟以及有关研究的不断深入, 提供了一种新的 Web 信息检索方法, 它智能化地将信息检索技术、知识学习技术有机地结合在一起, 构建了新的信息检索模式和策略检索机制。

2 Agent 技术

Agent 如今已成为人工智能、计算机等领域的一大研究热点^[1,5]。Agent 概念的提出及其应用研究成果, 使人们认识到 Agent 对于创建智能系统的意

义。Agent 在英语中是个多义词, 国内学者将其翻译为“主体”、“智能体”、“智能代理”等, 但这些译法都不能准确地反映 Agent 的本意。可以将 Agent 看作一个实体, Agent 通过它本身所固有传感器去感知外界环境, 通过其自身的执行器与外界环境发生作用。如图 1 所示。

Agent 是一个高度开放的智能系统, 其结构将直接影响到系统的智能和性能^[2,6]。Agent 的应用就是设计 Agent 程序, 即实现 Agent 从感知到动作的映射函数。这种 Agent 程序需要在某种结构的计算机上运行。简单的 Agent 结构可能只是一台计算机, 复杂的 Agent 结构可能包括用在特定任务上的特殊的硬件设备, 比如象图像采集设备或声音滤波设备等。Agent 结构还包括隔离纯硬件和 Agent 程

收稿日期: 2011-05-04; 修回日期: 2011-06-28

基金项目: 国家“十一五”科技支撑计划重点项目(2008BADA8B03)

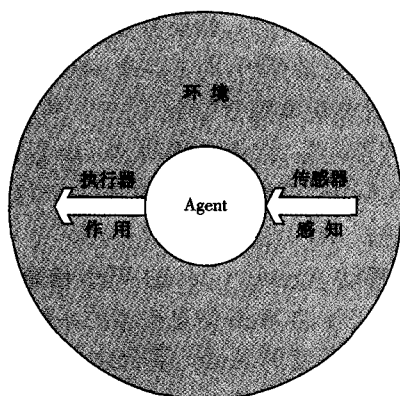


图1 Agent 与环境的交互作用

序软件平台,在它的基础上可以更高层次地进行编程。总之,Agent 结构接收传感器的输入,然后运行 Agent 程序,并将执行的结果传送到效应器进行动作。

一般被称为 Agent 的软件与硬件平台都具有以下特性:

(1)自主特性(Autonomy)。Agent 的自主特性表现在它能自主控制它本身的行为和状态,具有主动的、自发的特征,有目标和有意图地根据环境变化需求对其行为进行规划和设计。

(2)社交能力(Social Ability)。Agent 的社交能力主要表现在能与外界环境进行交互,它能感知外界环境,并运用自身的能力,对外界环境做出自身适当的反应^[3]。

(3)预动性(Pre-activeness)。其预动性表现在能通过 Agent 倡议去展示(规划和启动)有目标的行为,而非简单地适应环境变化而行事^[4]。

(4)反应能力(Reactivity)。Agent 的反应能力表现在当其位于某一特定的环境中,它能很快获取外部环境的状态信息、事件和行为,并对外界环境作出一定的反应来影响环境,与外界环境保持协调一致。从而与外界环境形成对立统一的关系,相互作用,互相依存^[7]。

3 传统的 Web 信息检索方法

Web 信息检索的主要任务是对 WWW 上的文档信息进行获取、表示、组织、存储和访问。传统的 Web 信息检索系统一般有索引数据库、查询服务、搜索引擎和 Robot 组成,其系统结构与组成如图 2 所示。

传统的 Web 信息检索的工作的基本原理是:检索系统过一段时间,它的搜索引擎就会分派一些

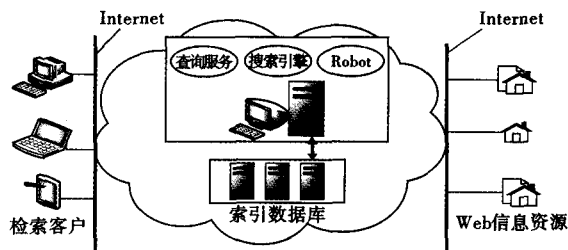


图2 传统的 Web 信息检索体系结构

Robot,首先从不同类别和被访问频率高的 URL (Uniform Resource Locator) 开始,获取 URL,并下载 URL 所指向的所有内容。此时,Robot 记录下该 URL 中所包含的 Web 超链,并按照其搜索算法开始对 URL 资源进行遍历,直遍历完成为止^[5]。搜索引擎会对获取的 URL 资源进行一定的预处理,并根据一些策略将其关键信息,如网页地址、标题、网站内容、关键字、摘要、特定字等存入索引数据库,建立本地查询数据库。这样,客户端进行查询时,主要访问本地的查询数据库,而不需要每次都重取 URL 资源。

Web 网站的内容更新很快,而传统的 Web 信息检索时,就受到了一定的局限性。Web 页面更新时,就要启动 Robot 开展本地查询数据库更新,而它驻留在 Robot 库中,不能迁移。它在 Internet 上通过 HTTP 与 Web 服务器进行通信时,对网络接的持续性提出了很高的要求,而且带宽开销也很大。当前 Web 页面基本都实现了动态管理,这使 Web 页面内容更新的频度增大,造成本地的查询数据库的更新较为缓慢,如果查询数据库越来越大,就会造成了查询数据库的更新不及时,不能反应地真实的 Web 内容。

4 基于移动 Agent-Aglet 的 Web 信息检索方法

随着 Agent 技术的成熟,可以运用跨平台语言,来解决传统的信息检索模式中网络传输问题,这为搜索模式提供了一种新思路和新方法。利用移动的 Agent 的检索模式取代传统的运用 Robot 检索模式,可以大大降低对网络连接持续性的要求和带宽的巨大开销,从而可以提高查询服务效率。

4.1 基于移动 Agent 的 Web 信息检索原理

基于移动 Agent 的 Web 信息检索的基本原理是:利用移动 Agent 来替代传统模式中的 Robot,根据移动 Agent 具有可迁移性的特点,它可以被派送异地的 Web 服务器来协助完成检索的任务,它在异

地 Web 服务器对 Web 页面内容进行对比,只送回更新的数据,这样有效地减少的网络开销,也有效保证了检索结果与现实的 Web 内容的一致性,变集中搜索模式为分布式信息查询模式^[7]. 其原理图如 3 所示. 工作原理为:

(1) 各 Web 站点首先会向索引服务器注册自己,注册的内容主要包括初始的 URL. 索引服务器会对注册的 Web 站点进行有效的分类.

(2) 查询服务在接收到用户的查询要求后,依据查询条件索引服务器开展查询,索引服务器返回给用户一系列关于特征匹配的 Web 站点,同时对移动 Agent 进行初始化,根据已制订的路由策略,向第一个 Web 站点服务器派送移动 Agent.

(3) 移动 Agent 在 Web 服务器上访问站点,进行识别和鉴定,再将符合要求的 Web 站点内容发送给用户.

(4) 当移动 Agent 完成任务后,它会根据约束

条件来决定下面的行为,如果需要再次转移,则会当前状态和历史记录存档,向下一个 Web 服务器发送迁移请求,请求获准则复制到下一个 Web 服务器,继续完成搜索任务,结束后移动 Agent 返回.

(5) 实际上,系统还可以发送多个移动 Agent 相互协作进行 Web 信息的检索.

4.2 基于移动 Agent-Aglet 的 Web 信息检索

基于 Agent 的 Web 信息检索的系统,就建立在 IBM 日本公司开发的 Aglet 平台之上,来实现基于 Agent 的 Web 信息检索.

Aglet 的系统框架如图 4 所示. Aglet 在运行时以下几个阶段:在 Aglet 往异地机器发送时,首先向 Aglet Runtime 层发出请求,Aglet Runtime 层会把 Aglet 发送的代码与状态信息进行序列化转换;其次,当其发送请求获准,系统会把转化后的代码与状态信息传送到 ATCI(Agent Transport & Communication Interface)层来进行处理,该层提供可用

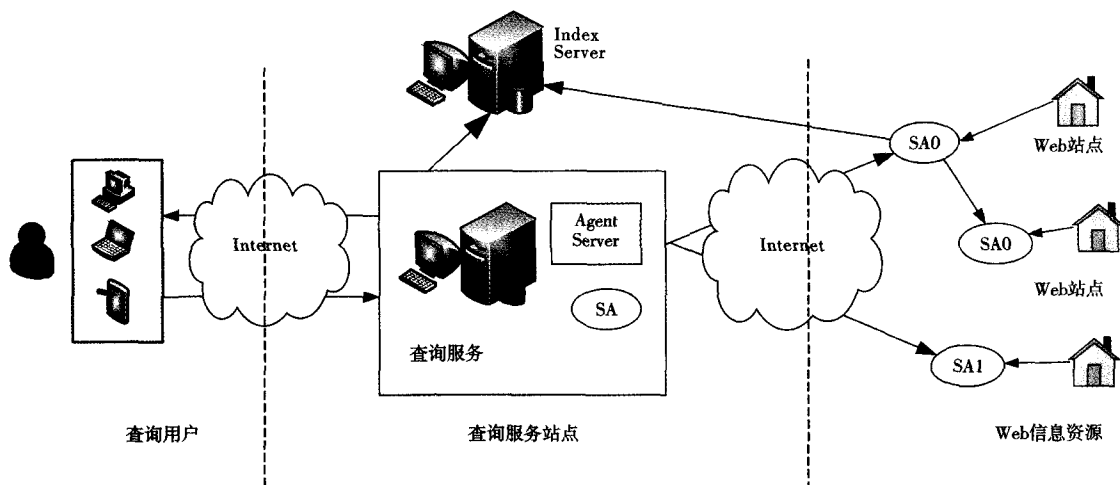


图 3 基于移动 Agent 的 Web 信息检索系统结构

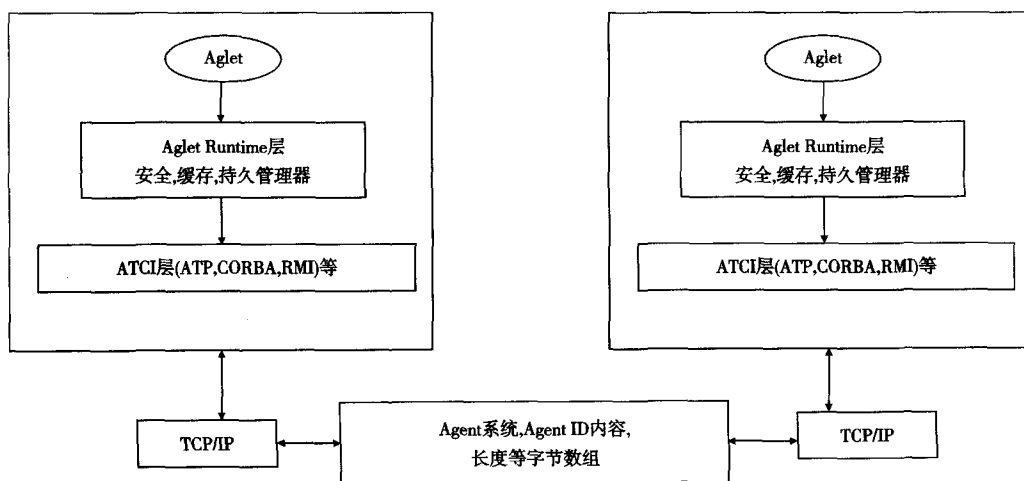


图 4 Aglet 平台的系统框架

的 ATP (Agent Transfer Protocol) 等接口。此后,系统会将转换后的数据附上系统信息,如系统名称、ID 等,以 bit 流方式传至异地主机;异地主机利用 ATCI 层所提供的 ATP 接口获取以上信息;最后,Aglet Runtime 层对以上信息进行反转换,获得 Aglet 的状态信息及代码,Aglet 即可在异地主机上运行。

Aglet 平台启动运行时,会提供一个 Context 来控制和管理 Aglet 的行为,如 Aglet 的创建、复制、派送、处理、收回等。Aglet 是通过消息传递方式进行通信的,考虑到安全性,Aglet 通过代理与外界沟通,只在 Local Machine 和 Remote Machine 之间进行信息交流,不直接处理诸如网络连接与通信等问题。

4.3 基于移动 Agent-aglet 的 SearchAglet 实现

由于移动 Agent 可以由索引服务器来收集 Web 站点注册,也可以脱离索引服务器,直接向远端 Web 站点派送移动 Agent,为了实现简便起见,SearchAglet 的实现,未使用索引服务器^[8]。

SearchAglet 在运用 Java 语言设计时分为三个模块:SA、AgletSever 和用户接口。SA 模块的主要功能是派送移动 Agent 到异地 Web 服务器来完成检索任务。移动 Agent 会根据路由策划,携带自身代码、身份验证信息及状态参数,从一个 Web 服务器迁移到另一个 Web 服务器,进行信息检索,并送回检索结果;AgletSever 的主要功能是负责对 SA 进行身份验证、提供驻留环境、以及底层通信机制。SA 只有通过验证后,才可以通过 Aglet 接口来访问本机的资源。在运行时,搜索主机和异地服务器主机上都要驻留 AgletSever;用户接口的主要功能接收并根据用户的查询条件,来选择的 Web 站点,并对 SA 进行初始化工作,为 SA 进行迁移路由策划,设置 SA 最大停留时间等参数。最后,用户接口将查询结果反馈给用户。

SearchAglet 依托 Aglet 平台,采用 Java 语言实现,主程序封装了三个类,即 InputDialog 类、OutputDialog 类和 SearchAglet 类,InputDialog 和 OutputDialog 主要完成查询条件的输入和检索结果的输出,SearchAglet 类实现了一个具有检索功能的 SA,来实现基于移动 Agent 对 Web 站点的信息检索。实质上,SearchAglet 就是一个具有检索功能的移动 Agent。

SearchAglet 在其实现时,主要涉及三个方面的关键技术,一是 SearchAglet 的方法驱动,二是迁移

路由策略的实现,三是完成检索功能。

(1) SearchAglet 方法驱动

SearchAglet 是基于 Aglet 平台,Aglet 平台提供了两种驱动方法,一是事件驱动,二是消息驱动。SearchAglet 主要靠消息驱动方法来实现通信。考虑到其安全性,移动 Agent 禁止外界直接获取其状态等信息,而是通过代理提供接口进行信息交流。其信息传递的方法为:

```
try{
    AgletProxy app2=InputDialog.agent.getProxy();
    App2.sendOnewayMessage(new Message("startTrip",
dest.getAddress()));
}catch (Exception e){
    e.printStackTrace();
}
```

(2) SearchAglet 迁移路由策略

Aglet 平台提供了两个类来完成移动 Agent 的迁移路由策略的设计与实现,即 seqItinerary 和 simpleItinerary。seqItinerary 复杂一些,该系统中使用了 simpleItinerary,它实现简单,只用对其对象 itinerary1 执行 itinerary1.go(destination,"###"),就可以将移动 Agent 派送到目标站点服务器,同时向移动 Agent 进行消息发送,驱动方法。这种消息传递属内部传递,不需要代理来完成。

(3) SearchAglet 对 Web 站点的信息检索

当 SearchAglet 将移动 Agent 派发到 Web 站点服务器后,移动 Agent 采用遍历的方法查询 Web 站点上的所有信息。程序中采用 URLs 队列来保存未访问的超链,用 VisitedURLs 保存已访问过的超链,用 Mactched 队列保存条件匹配的 URL。先指定 Web 站点的主页作为初始的 URL,将其放入 URLs 队列中,以后都从 URLs 中获取资源访问,直至为空。当取出一个地址,在 URLs 中进行消除,放入 VisitedURLs 中,同时获取资源,进行分析和条件匹配,如果满足条件匹配,将其存入 Mactched 队列中,这样,遍历完 Web 网站上的所有页面。最后,将 Mactched 队列中的 URL 返回给用户。

5 结束语

通过 SearchAglet 在 Agent-aglet 上的实现,虽然 SearchAglet 都采用了最简单的方式,但它证明了运用移动 Agent 实现对 Web 站点的信息查询的方法是可行的和有效的,为 Web 信息检索领域提供了一种新方法,将此方法应用到对 Internet 信息资

(下转第 176 页)

4.2 SIP 协议

SIP 是一个应用层的信令控制协议,用于创立、修改和释放一个或多个参与者的会话.会话的参与者可以通过组播、网状单播或两者的混合体进行通信.

本文设计的体系架构的应用点:用户设备(UE)通过 SIP 协议经过 IMS 后,通过 Presence 服务器或 Group 服务器读写 XML 数据.

4.3 SOAP 协议

SOAP(Simple Object Access Protocol)简单对象访问协议是一种简单的、轻量的、基于 XML 的协议,它是一个轻量级的消息传递协议,通过 SOAP 可实现 Web 服务间的消息传递,可在 WEB 上交换结构化的和类型化的固化信息.

本文设计的体系架构的应用点:Web Service 通过 SOAP 协议调用 RCS 服务资源.

通过 SOAP 协议实现调用 RCS 服务资源的能力,使的不具备 RCS 能力的用户设备,也可以通过 Web 的方式获得类似的 RCS 用户体验.

5 结束语

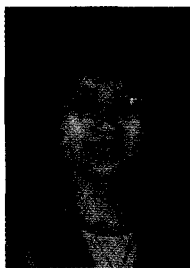
本文设计的 RCS 架构是在 XDMA 体系基础上构建的,符合 RCS 组织标准,并具有接入无关性和可扩展性好的特性,在呈现和群组服务上可容易地扩展诸如内容共享、即时消息等增值业务.并且,为向 RCS 平滑过渡提出了基于 SOAP 协议的无 RCS 能力设备的接入方式.对 RCS 富通信套件融合通讯

体验的应用及迅速发展具有现实意义.

参考文献:

- [1] 彭青立,黄海,朱春梅.移动互联网时代的通信利器——RCS终端[J]. 电信技术,2011(02):33-36.
- [2] RCS R1-R3 版. [EB/OL]. [2011-05-05]. http://www.gsmworld.com/our-work/mobile_lifestyle/rcs/rcs_specification_documents.htm.
- [3] Open Mobile AllianceTM XML Document Management Architecture,OMA-AD-XDM-V2.0. [EB/OL]. [2011-05-05]. <http://www.openmobilealliance.org/>.
- [4] 冉崇善,米微琪,史宇. XDMS:基于 XML 的数据管理系统[J]. 微电子学与计算机,2009(03):166-168.
- [5] 程佳鸣,糜正琨,徐名海,等. 融合地址簿系统的订阅实现技术[J]. 移动通信,2010(10):93-97.

作者简介:



张小芳 女,(1971—),博士,副教授.研究方向为系统架构、软件工程、集群操作系统、作业管理系统.

梁凯 男,(1986—),硕士研究生.研究方向为电信平台、数据库技术的研究、集群管理系统.

王会峰 男,(1981—),高级工程师.研究方向为电信业务及应用研究.

(上接第 171 页)

源的数据挖掘上,能够有效地节约网络带宽开销,提高查询效率.

参考文献:

- [1] 张鹏飞. Agent 通信的全信息形式语用方法[D]. 北京:北京邮电大学,2007.
- [2] 吴铁洲,贲可荣. 移动 Agent 技术及实现方法[J]. 燕山大学学报,2005,29(4):332-336.
- [3] 李国勇,李维民. 人工智能及其应用[M]. 北京:电子工业出版社,2009.
- [4] 贾松浩,杨彩,张海玉,等. 基于本体的 Web 改进的信息检索方法[J]. 微计算机信息,2009(33):87-89.
- [5] 张建芬. 基于移动 Agent 的图书馆信息检索系统设计[J]. 图书馆学研究,2010,(01):18-20.
- [6] 崔亚楠,黄文明,朱英. 基于 Agent 的网格资源共享方法研究[J]. 微电子学与计算机,2009,26(12):112-116.

[7] 姜艺,张辉,朱俊武. 一种基于本体的多 Agent 语义挖掘模型[J]. 微电子学与计算机,2009,26(10):189-192.

[8] 李贺华,林婧,单光庆,等. 基于 Agent 的 Web 信息检索应用研究[J]. 通讯技术,2009,(08):051-053.

作者简介:



徐朝辉 (1970—),副教授.研究方向为计算机网络、数据挖掘.

秦杰 (1968—),博士,副教授.研究方向为计算机网络.

甄彤 (1965—),教授,硕士生导师.研究方向为计算机应用.