

一个个性化的信息搜集 Agent 的设计与实现*

潘金贵, 胡学联, 李俊, 张灵玲

(南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093);

(南京大学 多媒体计算技术研究所, 江苏 南京 210093)

E-mail: panjg@nju.edu.cn; lijun@mes.nju.edu.cn

http://mes.nju.edu.cn

摘要: 介绍了一个个性化的信息搜集 Agent 系统——DOLTRI-Agent (distance and open learning training resource information agent) 的设计与实现。DOLTRI-Agent 系统是一个远程开放式学习 (distance and open learning, 简称 DOL) 的教材信息搜集 Agent, 它基于多 Agent 体系结构, 包含 3 个子 Agent 和其他 3 个组件。该系统具有主动搜集、维护、更新信息的功能, 同时还具有个性化的特点, 可不断学习用户知识, 包括用户感兴趣的信息领域、用户的使用习惯等, 并能运用这些知识来为用户提供个性化的信息服务和操作界面。

关键词: Agent; 信息搜集 Agent; 个性化 Agent; 多 Agent 系统; 远程开放式学习

中图法分类号: TP393 **文献标识码:** A

Internet 尤其是 WWW 的飞速发展, 给人们带来了前所未有的信息共享与交流。但是, Internet 上信息的特点导致了人们在越来越依赖它的同时面临着在庞大的信息空间中迷失方向的困惑。而且, 随着 Internet 上信息的急剧增长, 现有的搜索引擎所能提供的功能也难以满足人们的需要。由于 Agent 技术的诸多特性, 在信息搜索中引入 Agent 技术可以使搜索引擎的缺陷不再成为人们有效利用网络资源的瓶颈, 利用 Agent 帮助查找用户所需信息使得 Internet 信息空间问题的解决“柳暗花明又一村”。

Internet 的发展为远程开放教育提供了极好的机遇和基础。远程开放教育是教育走向全球化的一个有效途径, 实施远程开放教育的一个重要任务是设计和开发多媒体化的教学课件。DOLTRI-Agent (distance and open learning training resource information agent) 系统是我们研究开发的 Agent 技术在 Internet 信息领域中的应用。它为远程开放教育的教材设计者提供了一个个性化的能够自主地在 Internet 上搜集教材信息的工具。

1 DOLTRI-Agent 系统的体系结构和工作原理

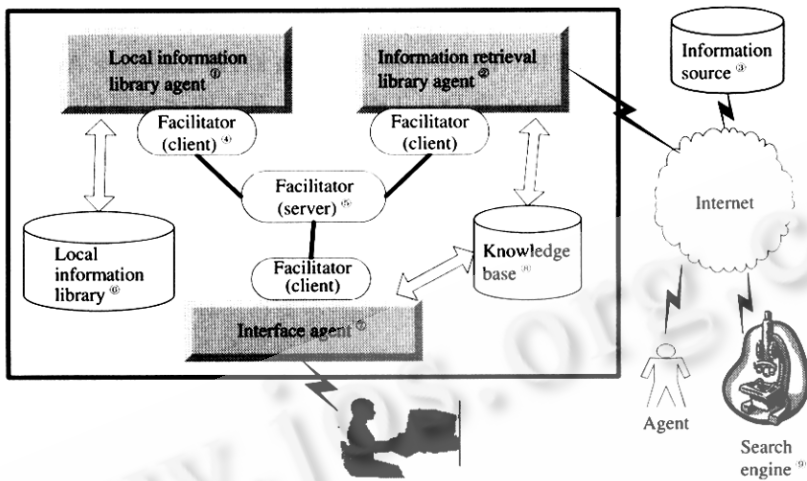
DOLTRI-Agent 系统基于多 Agent 的体系结构, 由 3 个子 Agent——界面 Agent、本地信息库 Agent 和信息搜集 Agent 以及一个本地信息库、一个知识库和一个 Facilitator 组成, 它们互相协作, 共同完成对用户所需信息进行检索、收集和维持的任务。它们之间的关系以及与环境的关系如图 1 所示, 其中黑框内的部分为 DOLTRI-Agent 系统的体系结构。

* 收稿日期: 2000-06-20; 修改日期: 2001-02-16

基金项目: 日本邮政省通信广播机构(TAO)国际合作研究基金资助(861)

作者简介: 潘金贵(1952—), 男, 江苏丹徒人, 教授, 博士生导师, 主要研究领域为中间件, Agent 技术, 多媒体远程教育; 胡学联(1952—), 男, 安徽黄山人, 副教授, 主要研究领域为 Agent 技术, 软件工程; 李俊(1977—), 男, 江苏武进人, 助教, 主要研究领域为智能信息收集, Agent 技术, 无线接入技术, 无线网络应用; 张灵玲(1975—), 女, 江苏武进人, 硕士, 主要研究领域为信息搜索技术, Agent 技术。

下面分别介绍该系统的各个组成部分的设计与实现.



- ①本地信息库 Agent, ②信息搜集 Agent, ③信息源, ④通信协调机构(客户端),
⑤通信协议机构(服务器端), ⑥本地信息库, ⑦界面 Agent, ⑧知识库, ⑨搜索引擎.

Fig. 1 Architecture and environment of DOLTRI-Agent

图 1 DOLTRI-Agent 系统的体系结构及其与所处环境的关系图

1.1 界面 Agent^[1]

界面 Agent (interface agent, 简称 IA) 的主要功能是接受用户的查询请求, 主动帮助用户细化查询要求, 并显示查询结果, 同时在这个过程中学习并记录用户的兴趣和习惯. 它还负责显示本地信息库的内容, 并在界面中反映本地信息库的动态变化. 界面 Agent 主要由用户界面、与其他 Agent 的接口部分、分析推理机、使用记录数据库这 4 个部分组成. 其中, 用户界面即通常的 GUI; 与其他 Agent 的接口部分定义了一些 Agent 之间通信和交互的规则与方法; 使用记录数据库来记录用户的使用历史; 分析推理机是 IA 的大脑, 它包括很多用于分析推理的算法和规则, 并需要用到相关的一些知识.

1.2 信息搜集 Agent

信息搜集 Agent (information retrieval agent, 简称 IRA) 是由信息任务模块 (information task module, 简称 ITM)、信息搜索模块 (information search module, 简称 ISM) 和信息分析模块 (information analysis module, 简称 IAM) 这 3 个功能模块组成的. 信息任务模块 ITM 从 IA 处接受用户提出的查询请求, 生成搜集信息的多个子目标提交给信息搜索模块 ISM. ISM 访问网上异构的信息资源, 完成在 Internet 上搜集教材信息的任务. 目前, 信息搜索模块采用的信息搜索方式是与一些知名的信息搜索引擎和信息服务提供者 (如 Yahoo, Altavista, Infoseek 等) 合作. ISA 将得到的结果再交给信息分析模块 IAM, 以便对合作的信息提供者返回的结果信息进行再分析. 运用系统的知识对搜索结果进行过滤, 再通过 IA 返回给用户.

1.3 本地信息库 Agent

本地信息库 Agent (local information library agent, 简称 LILA) 主要由用户信息管理模块 TRILA、教材信息搜索模块 TRILSA 和教材信息实时更新模块 TRILRTUA 这 3 部分构成.

1.4 Facilitator

DOLTRI-Agent 系统基于多 Agent 体系结构. 该体系结构内的各个 Agent 可以分布在网络的多个计算机节点上运行, 而且 DOLTRI-Agent 支持多用户查询, 即 Interface Agent 可以在网络的多个计算机节点上复制运行. Facilitator 为 DOLTRI-Agent 系统的运行提供了环境. 它主要有以下功能: 通信服务、传输服务和安全机制. 目前, Facilitator 只负责通信任务, 使用 KQML 作为通信语言^[2].

1.5 知识库

知识库内存放了用户知识(user knowledge)和领域知识(domain knowledge). 它在整个系统中占有十分重要的地位. 在信息搜集任务的执行过程中, 实现智能性和提高搜索精度的 3 个关键阶段——查询处理、信息再分析和相关性反馈都要与知识库进行交互与合作. 知识库由以下 4 个部分组成: 通用字典(stopwords dictionary)、关键字字典(keywords dictionary)、通用分类模型(general classification model)和用户模型(user model).

1.6 本地信息库

本地信息库(local information library, 简称 LIL)建立在 Oracle 数据库上, 因此可以直接使用其强大的数据维护和管理工具. 本地信息库用来存储用户满意并需要保存的信息. 信息按主题存放, 这样, 搜索起来效率很高. 当然, 用户也可以按其他属性查询. 在本地信息库中存储了文档的元信息(meta data)(如 URL、作者、标题、日期、关键字、摘要、有效期等)以及存储该文档的用户 ID. 由于本系统是多用户的, 各用户可以分别存储自己感兴趣的信息, 也可将其集中存放在一张表中. 在查询本地信息库时, 也可指定范围为该用户存放的信息或共享所有信息.

2 系统特性

DOLTRI-Agent 系统具有以下特点: 个性化、自主性、信息维护功能、知识维护功能、协作性、灵活性与可扩展性. 下面我们详细介绍该系统的个性化与自主性, 并简单介绍其他特点.

2.1 个性化

DOLTRI-Agent 系统的个性化特点是指能够满足用户的个性化需求. 它的界面 Agent 能够不断地从与用户的交互和反馈中学习到用户知识, 包括用户的兴趣、偏好、习惯、思维方式等, 因此, 在对用户提供服务时能更好地提高效率并体现用户的特性, 让用户满意. DOLTRI-Agent 系统的界面 Agent 学习用户知识的方式有 3 种: 观察用户的操作、接受用户的反馈信息和接受用户的直接训练. DOLTRI-Agent 系统利用学习到的用户知识为用户提供的个性化服务主要表现在查询请求的细化和文档信息的分析这两个方面.

2.2 自主性

自主性是指 Agent 系统能自行安排和执行某个任务, 不需要用户的指导^[3]. 因此, Agent 具有各自的内在的状态和目标, 它们的行为是代表其用户达到某个目标. DOLTRI-Agent 系统的自主性表现在两个方面: 首先, 系统定期地主动更新本地信息库中的信息; 其次, 系统在获得一定程度的用户知识以后, 能预测用户的行为并自主地为用户作出信息选择和决定, 帮助用户处理搜索结果, 而不需要与用户一步步地进行交互.

2.3 其他特点

DOLTRI-Agent 的其他特点还包括信息维护功能、知识维护功能、协作性、灵活性与可扩充性。

3 运行结果

我们于 1999 年底完成了 DOLTRI-Agent 系统的开发工作,2000 年开始将其投入试用并进行了功能的改进. 南京大学计算机科学与技术系的部分教师使用该系统之后,认为得效果不错. 该系统相对于一般的搜索引擎而言,搜索的精度大大提高了. 特别是,经过一段时间的学习之后,系统能够提供一些智能性的即时帮助、主动操作等等,大大减轻了用户的负担. 其主要的不足在于,刚开始使用的时候不能提供帮助,需要很多额外的交互;另外,在用户转变兴趣时也会产生这样的问题,并且有时候还会提供错误的提示等等. 下面给出一个例子.

搜索请求:software agent(关键字).

请求的细化:software agent: software and operating system ->Computers & Information Technology.

搜索的结果:结果数量可以由用户指定,系统会列表显示这些结果(如图 2 所示),用户可以通过浏览器浏览某个结果的详细内容(如图 3 所示). 同时,系统会根据自己的知识对结果进行预处理,用户也可以进行反馈.

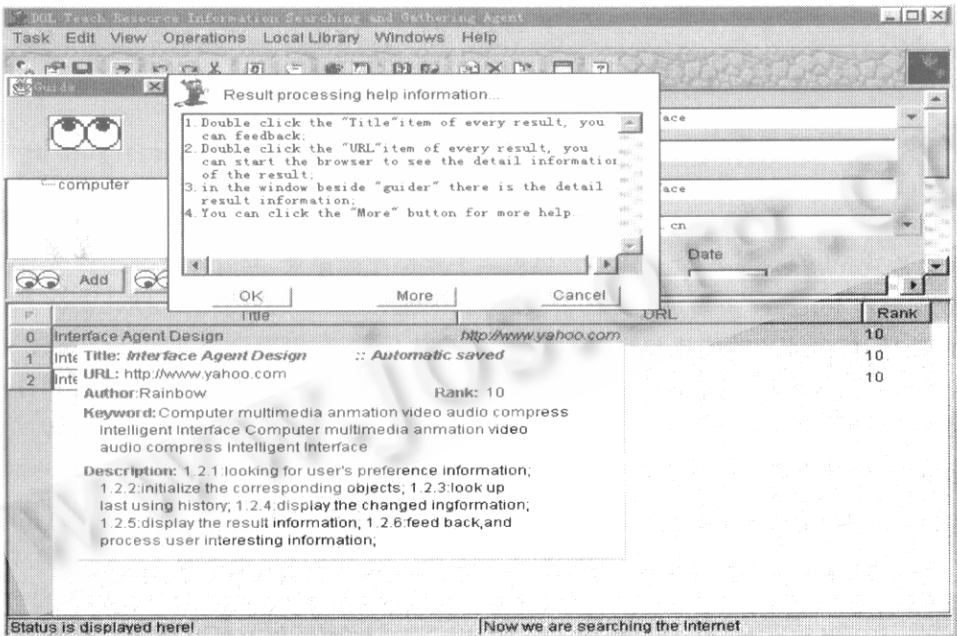


Fig. 2 Results from Internet

图 2 网络搜索结果

4 与相关工作的比较

国际上与信息 Agent 相关的比较有名的系统有:美国华盛顿大学(UW)的 Internet Softbot 系

统、美国南加州大学(USC)信息科学研究所(ISI)的 SIMS 系统、瑞典 SICS 实验室的 PUSH(plan-and user sensitive help)系统、美国 Stanford 大学的 Fab;adaptive Web page Recommendation service,等等。下面我们概要介绍 Internet Softbot 系统和 SIMS 系统,并与 DOLTRI-Agent 系统进行比较。

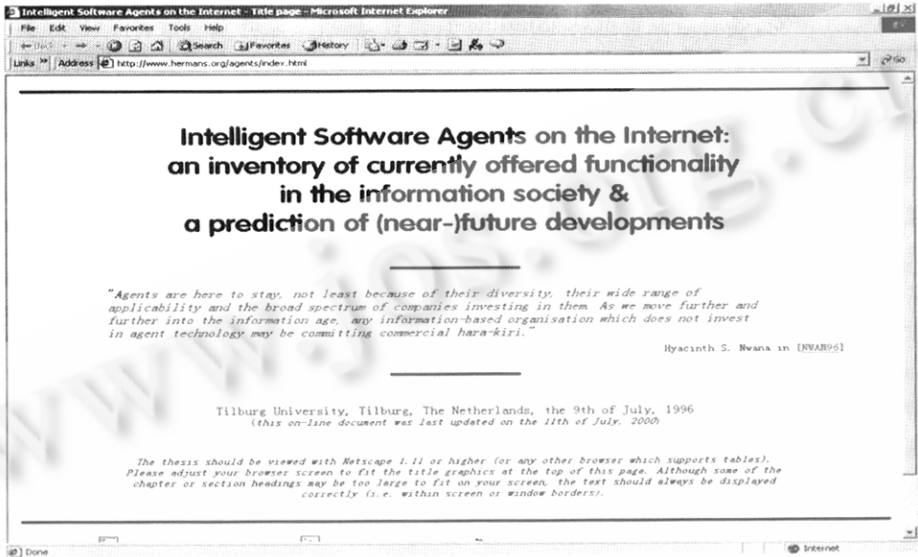


图3 详细的结果信息

Fig.3 Detail information of one result

4.1 Internet Softbot 系统^[3]

Internet Softbot 系统是由华盛顿大学(University of Washington)的 Etzioni 和 Weld 在 1994 年开发的一个信息 Agent 的实例。Nwana 认为该系统是目前信息 Agent 最新发展的代表^[4]。Softbot 的含义是 software robot。它允许用户提出一个高层次的请求,并且能够利用搜索和推论的知识来决定如何在 Internet 上满足用户的请求。与 Internet Softbot 系统相比,DOLTRI-Agent 系统的不足之处是:信息搜索功能比较简单;网络功能也比较单一,没有容错功能。DOLTRI-Agent 系统优于 Internet Softbot 系统之处主要是充分运用了 Agent 技术,系统具有学习能力和自动维护动态信息的功能。

4.2 SIMS 系统^[5]

SIMS 系统是南加州大学(University of Southern California)信息科学研究所(Information Science Institute)的 Knoblock 和 Ambite 深入开发的一个用于信息收集的 Agent 实例。DOLTRI-Agent 系统与 SIMS 系统的功能比较类似,都拥有较为详细的领域知识,具有模块性和可扩充性,通过 KQML^[2]传递消息等。但比较而言,DOLTRI-Agent 系统的优点是能够学习获取并充分使用用户知识,对于不同的用户,系统能根据不同的侧重点去使用信息分类库;同时,该系统的界面 Agent 采用了对多场景支持的技术,使用户在教材信息搜集的各个步骤中都能得到智能的支持服务,大大减轻了用户使用本软件的培训负担。DOLTRI-Agent 系统与 SIMS 系统相比,其不足之处有:(1)缺乏不同信息源的模型;(2)缺少对现有数据库和应用程序进行包装的功能;(3)缺少追踪不同 Agent 的语义的功能。

5 总结与进一步的工作

DOLTRI-Agent 系统是一个基于 Internet 的信息搜集 Agent 系统. 它能为远程开放式教学的教材设计者以及其他用户提供个性化的信息搜集服务. DOLTRI-Agent 系统具有主动搜集、维护、更新信息的功能;同时,它还具有个性化的特点,能不断学习用户知识,包括用户感兴趣的信息领域、用户的使用习惯等,并能运用这些知识来为用户提供个性化的信息服务和操作界面.

鉴于 DOLTRI-Agent 系统与相关系统比较的不足,今后的研究工作主要包括:(1) 建立更多信息源的模型. 目前本系统仅与搜索引擎合作搜集信息,将来会扩展到其他信息源,如电子图书馆以及其他信息 Agent 系统等;(2) 引进移动 Agent 技术. Agent 移动到信息源上分析信息,仅传回相关信息的元信息,这将大大节省网络传输带宽;(3) 扩充搜集信息的语言种类. 目前仅限于英语,不久将扩充中文;(4) 不断改进信息分析算法和学习算法.

References:

- [1] Li, Jun, Zhang, Ling-ling, Zhou, Wen-hui, *et al.* Design and implementation of an intelligent user interface agent. *Journal of Software*, 1999, 10(8): 835~842 (in Chinese).
- [2] Finin, T., Labrou, Y., Mayfield, J. KQML as an agent communication language. In: Bradshaw, J. M., ed. *Software Agents*. Cambridge, MA: MIT Press, 1997.
- [3] Nwana, H. S. An overview, knowledge engineering review. *Software Agents*, 1996, 11(3): 205~244.
- [4] Nwana, H. S., Ndumu, D. T. A perspective on software agents research. *The Knowledge Engineering Review*, 1999, 14(2): 1~18.
- [5] Bradshaw, J. M. An introduction to software agents. In: Bradshaw, J. M., ed. *Software Agents*. Cambridge, MA: MIT Press, 1997.

附中文参考文献:

- [1] 李俊,张灵玲,周文辉,等. 一个智能用户接口 Agent 设计与实现. *软件学报*, 1999, 10(8): 835~842.

Design and Implementation of a Personalized Information Retrieval Agent*

PAN Jin-gui, HU Xue-lian, LI Jun, ZHANG Ling-ling

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China);

(Institute of Multimedia Computing Technology, Nanjing University, Nanjing 210093, China)

E-mail: panjg@nju.edu.cn; lijun@mes.nju.edu.cn

http://mes.nju.edu.cn

Abstract: The design and implementation of a personalized information retrieval agent, DOLTRI-Agent (distance and open learning training resource information agent), is introduced in this paper. DOLTRI-Agent is to retrieve training resource information for DOL (distance and open learning). It is based on the multi-agent architecture, including three sub-agents and three other components. DOLTRI-Agent system owns the autonomous functions of information retrieval, maintenance and updating. Furthermore, it has the feature of personalization: it can continuously learn user knowledge, such as user interested information fields and user's habits; can also utilize this knowledge to provide the personalized information services and interactive interface.

Key words: agent; information retrieval agent; personalized agent; multi-agent system; distance and open learning

* Received June 20, 2000; accepted February 16, 2001

Supported by the Research Foundation of Telecommunications Advanced Organization (TAO) of Ministry of Posts and Telecommunication of Japan under Grant No. 861