



面向网络教育的个性化智能搜索引擎的设计与实现

内容编辑: 刘锴 / 网上发布: 2007-9-19 / 已经查看: 10927次

吴丽华 冯建平 罗云锋

【摘要】 本文以当前网络教育资源环境为背景, 论述了智能代理和智能搜索代理技术的基本概念、特点和研究现状, 给出了一个基于智能搜索代理技术的个性化智能搜索引擎的设计及实现方法。其目的在于, 探究一种能为学生提供个性化信息服务的智能代理系统, 同时也是对网络教育中资源搜索工具的智能化进行一次有益的尝试和探索。

【关键词】 智能搜索代理; 个性化; 概念网; 智能兴趣代理

前言

进入21世纪, 先进的计算机技术和网络通信技术正以前所未有的动力推动着人类社会各方面的进步, 尤其是Internet的迅速普及, 深刻地改变了人们的生活、学习和思维方式。但是, 由于Internet是一个开放、分布的信息空间, 网上资源以指数速度增长, 用户进行信息检索经常会出现“信息过载”和“资源迷向”。现有的搜索引擎已越来越难以满足人们高质量地获取网络信息的需求。因此, 帮助Internet用户根据个人的兴趣, 自动查找信息, 屏蔽不相关信息, 即提供网络环境下的个性化信息服务, 已成为当前网络信息检索的重要课题。近年来, 出现了许多满足用户个性化信息需求的技术, 如各类搜索引擎(垂直搜索引擎、主题网站等)、数据推送技术、过程跟踪技术、智能搜索代理和协同过滤等。其中智能搜索代理技术克服了传统搜索引擎的缺陷, 智能化地理解用户的信息需求, 以实现个性化的信息服务。

当前, 网络教育为学生提供了一个基于Internet的虚拟学习环境, 同样也给学生带来了一定的信息盲目性。本文以当前网络教育资源环境为背景, 论述了智能代理和智能搜索代理技术的基本概念、特点和应用现状, 并基于概念语义网络和智能搜索代理技术, 给出了适合网络教育特定领域的个性化智能代理搜索引擎的设计 and 实现方法。其目的在于: 探究一种能为网络教育资源环境下的学生提供个性化信息服务的智能代理系统, 同时也是对网络教育中资源搜索工具的智能化进行一次有益的尝试和探索。

智能代理技术

1. 智能代理



- [理论探讨] [中国高校教育技术学科综合竞 ...](#)
- [新闻快报] [中国教育技术协会2008年征文通知](#)
- [研究生教育] [教育技术学硕士研究生招生变 ...](#)
- [资源共享] [CSSCI来源期刊\(2008—2009年\)](#)
- [新闻快报] [第二届国际信息技术研讨会\(...](#)
- [专家学者] [汪琼 教授](#)
- [专家学者] [祝智庭 教授](#)
- [就业展望] [徐州师范大学2008年人才招聘](#)
- [课题奖项] [全国教育科学“十一五”规划 ...](#)
- [教育技术史] [思辨中演进的教育技术学\(上\)](#)

- [网络学习平台功能模块简析](#)
- [基于内容的图像网络教学资源检索研究](#)
- [从默会知识的视角看社会性软件在教师专业发展中的作用](#)
- [解析社会性软件及其在远程教育中的应用](#)
- [聚合社会性网络档案的20种方法](#)
- [超越用户创建的模式: WEB2.0和语义网络](#)
- [社会性软件: 学习的黄金之道](#)
- [高校课程管理系统的选择策略研究](#)
- [基于绩效观念的远程学习支持服务系统的设计](#)
- [如何正确引导远程开放教育学生的学习心理](#)
- [基于CORBA/ XML/ JAVA ...](#)
- [评论: 如何给明天的远 ...](#)
- [美国普通高校的远程教 ...](#)
- [教学软件开发流程优化 ...](#)
- [手持式网络学习系统在 ...](#)

智能代理技术(Intelligent Agent Technology)是分布式人工智能研究和开发的一个新领域,最早于20世纪90年代提出,最近几年发展迅速。其内容涉及人工智能、信息检索、计算机网络、数据库、数据挖掘、自然语言处理等领域。智能代理又称为智能体,实际上是一种软件单元,具有高度智能性和自主学习性。它可以根据用户定义的准则,主动地通过智能化代理服务器为用户搜集最感兴趣的信息,然后利用代理通信协议把加工过的信息按时推送给用户,并能推测用户的意图,自主制订、调整和执行工作计划。

伴随计算机技术和人工智能研究的不断发展,具有智能性、代理性、适应性、学习性等特征的智能代理技术逐渐走向实际应用,目前主要应用在智能搜索代理、数字图书馆、电子商务和远程教育等领域。

2. 研究现状及应用

智能代理可以作为虚拟的教师、虚拟的学习伙伴、虚拟的实验室设备、虚拟的图书馆管理员等出现在网络教育系统中,它增强了教学内容的趣味性和人性化色彩,改善了教学效果。在基于Internet的远程教育中,智能代理技术正逐渐取代ICAI而成为教学领域实现智能化的一种主流技术。它不仅可以作为“教师代理”,也可以作为“学生代理”,而且还可以成为学生学习过程中的“多重代理”。利用智能代理技术可以构造各种虚拟现实,从虚拟的人物到虚拟的社区,极大地丰富了远程教育的教学手段。

通常,智能代理技术在网络教育中的五个方面发挥着重要作用:实时监督、教学分析、信息检索、协作学习和智能推理。国内外这方面的研究及产品已趋于成熟,并已有投入实际应用的模型和系统。

Microsoft Agent(微软称之为“Office助手”)是大家最熟悉的一种“助手代理”,它最早出现在Office95中,主要用于Office的帮助。Adele(远距离教育代理简易版,Agent for Distance Education-Light Editions)是由美国南加利福尼亚大学开发的另一种典型的“教育代理”,该系统结构中最主要的部分是代理人和推理引擎,其中“代理人”是用Java语言编写的,可以保证平台的独立性和扩展性;推理引擎能够完成所有的监控和决策。Adele是智能代理的一项成功应用,它通过一个能够支持与学生进行连续和多种模式交互的虚拟“代理人”来完成最基本的教育功能,即表达知识、监控学生、提供反馈、探究问题,以及提示和解答。目前,它已应用于医学教育课件,教授外伤和肿瘤医疗课程,是一种专为适应虚拟教育环境而设计的教育代理。

国内这方面的研究主要集中在基于智能代理的信息搜索及代理间的协作方面,有些学者提出把智能代理技术应用在网络教学的智能信息检索中,学习者Agent通过用户界面和知识库管理系统来初始化自己的兴趣指向,这些兴趣指向被存储在知识库中的用户模型部分,然后由信息搜索Agent读取知识库中的用户模型,确定搜索任务,启动搜索引擎到某几个网站或整个Internet网上进行搜索。有些学者又提出一种把学习任务细化的思想,由各代理和相应的学习成员完成各项子任务,各代理成员之间通过KQML通信,共享彼此的研究成果。

智能搜索代理技术

1. 智能搜索代理

智能搜索代理(Intelligent Retrieval Agent)是智能代理技术在网络信息检索特定领域中的应用,它对用户信息需求、偏好进行甄别、归纳、总结,分析用户的兴趣,并借助学习的规则,自动、独立地代理用户查找信息,是目前具有前瞻性的网络信息检索工具。目前,智能搜索代理已经成为Web网络信息检索的核心技术。同传统的搜索引擎对比,其特色主要体现在以下几个方面:

(1) 网络信息收集的智能化:采取最为有效的搜索策略,按一定的语法规则智能地、有选择地自动收集网络信息。

(2) 网络信息处理的智能化:运用推理机制和学习机制,处理和理解收集来的网络信息智能。

(3) 网络信息检索的智能化:采用自然语言检索入口,允许用户自由表达查询请求。采用语义网络等智能技术,通过汉语切词、句法分析和统计理论有效地理解用户的请求。借助知识库和规则库中对用户行为和需求的描述规则,参考用户以前的需求记录和爱好,推断用户的最大可能需求。

(4) 网络信息检索的个性化:采用机器学习、用户行为建模、推理机制、规则描述等技术,通过学习了解用户的行为、爱好、兴趣,推理用户的潜在需求,可以根据用户的评价和反馈调整自己的行为,动态地关注用户所需信息的变化,实时地把最新信息推送给用户,实现服务的个性化。

2. 研究现状及应用

目前,智能搜索代理在网络信息搜索中的应用十分广泛。如Browser Buddy就是一个成功的智能代理系统,它是一个用于组织和链入Web页面的基于规则的智能代理,经过整夜的信息搜索,会在每天早上及时给用户提供一个服务清单。Autonomy智能代理是一个典型的“学习代理”(Learning Agent),它使用神经网络而不是关键词来识别信息的模式,并且在用户反馈信息的教导下不断训练,直至能有效地找到用户感兴趣的文档。

还有一些其它专门智能代理软件用来帮助用户寻找特定信息。如Firely等,使用信息过滤技术来帮助用户查询喜爱的电影和音乐;还有CMU的智能搜索代理WeDoggie,这是一个基于规则的系统,可以根据用户的兴趣推荐网络资源;CMU的智能新闻阅读器News Weeder,基于机器学习的理论来学习用户的兴趣,查找新的页面和文档。

个性化智能搜索引擎

1. 现有系统的缺陷

目前,网络教育中提供的各类搜索引擎往往适用于短暂的随机性查询,一般利用学生提供的关键词或查询条件搜索信息,对查询结果的排序算法主要依据关键词的词频、位置、邻近度、更新日期等指标,计算出各网页的相关度及权重,然后根据“相关度”的大小,按递减顺序将这些系统认为和查询相关的文档(网页链接)返回给用户。其局限性表现在:

(1) 简单的关键词匹配，往往输出大量的文档，而真正和用户信息需求相关的文本却很少，这使得学生耗费更多的时间和精力处理一些毫不相关的文本。

(2) 基于关键字的需求模型不能全面地反映用户兴趣，更没有用户兴趣模型保存和维护的功能。而用户查询与文档“相关度”的评价，是一种依据查询者兴趣趋向进行评价的主观评价方法。

尤其是针对网络教育资源环境下学习的学生，他们的兴趣往往表现在就业、考研、各类等级考试、专业电子期刊等方面，但目前常用的搜索引擎不能更准确、及时、权威地检索出他们需求的信息。因此，分析学生的兴趣类型或专业信息需求，建构学生的个性化兴趣模型，对搜索引擎所返回的结果进行过滤，以实现搜索引擎为学生提供真正意义上“所得即所需”的个性化信息资源服务是非常必要的。

2. 系统结构

基于上述思想，我们设计了一个基于网络教育资源环境的个性化智能搜索引擎，主要面向网络学习者。该智能搜索引擎分为四个部分：搜索Robot、索引数据库、信息检索模块和个性化智能兴趣代理，如图1所示。

系统中各模块实现的功能：

(1) 搜索Robot：按照一定的策略在网络教育资源中抓取网页，并将网页交给索引数据库进行存储。

(2) 索引数据库：采用全文检索技术，对搜索来的网页的全部内容进行基于“词”的索引，再对应概念语义网中所出现的“领域词”，用“词频法”计算出领域词在网页中出现的频率以表示该领域词与网页的相关度，最后按照词频大小进行排序，并形成倒排文档，存储在索引数据库中。

(3) 信息检索模块：进行查询子句的分词处理后，提取关键词，采用“概念树”结构，实现关键词的概念检索，将基于词的检索提高到概念层次。

(4) 个性化智能兴趣代理：通过对学生检索结果的个性化信息的提取，建构学生的个性化兴趣模型，并对模型进行维护和更新，以为学生提供个性化服务。然后利用数据挖掘技术对用户访问的历史信息进行兴趣规则抽取，以此来预测用户将来的行为。

3. 系统设计及实现

对于本系统，我们主要是从“信息检索模块”和“个性化智能兴趣代理”两个方面进行智能化研究。主要设计思想为：在全文检索的基础上，运用语义网络构建“概念网”，实现概念扩展，提高系统的查全率；再通过智能搜索代理建立学生的个性化兴趣模型，以滤出学生所需信息资源，提高系统的查准率。

(1) 概念语义网络的构建

“概念”(Concept)是在客观事物的基础上概括而成的,是客观事物在头脑中的抽象反映,它要通过字、词、词组等概念描述元素才能表达出来。“语义网络”是知识的一种图解表示,它是由节点和弧线或链线组成的。节点用于表示概念,弧线用于表示节点之间的关系。

“概念语义网络”的构建需要具有一定规模的丰富的知识作为基础,而且知识表达要准确、清晰,这样构建的概念网整体结构才能完整。而在网络教育环境中,领域知识的分类相对稳定而且准确,因此概念网的构建可以此为基础,其结构如图2所示。

概念网构建的方法如下:

①用“概念树”的方法建立概念之间的上下层关系。上层概念是其所有下层概念共同属性的概括,下层概念则是从不同角度对其上层概念加以细化。最上层是一个虚拟层,使整个概念树形成一个整体。

在这里,每个概念节点都可以按学科分类代码(参照1992年国家颁布的《中华人民共和国学科分类与代码国家标准》)为基础进行概念编码标识,并且每个概念都带有一个集合,是该概念的同义但不同描述元素组成的集合,比如: $\Phi(\text{计算机软件})=\{\text{软件, 程序, software}\}$ 。集合可以根据同义词典或实际需要进行添加、删除、修改等操作,这样处理还可忽略概念的语种差异,识别文档中存在的中英文互用。将这些信息存入概念库中,概念标识可表示为:

Code[Concept] {Descriptor1, Descriptor2, ..., Descriptorn}

②每个概念可与其它概念建立相应的关系。这种关系是不同于分类中上下层关系的横向关系,可采用不同的弧来表达概念之间的不同关系。

(2) 基于概念的检索系统

本系统的“分词词典”中提供了一个主词典、一个同义词典及一个蕴涵词典。当学生输入查询请求后,由“分词词典”进行分词,提取领域词,并根据概念语义网赋以相应的概念编码标注,检索模块将对其进行概念的扩展。具体地说,实现了概念层次上的同义扩展检索、概念拓展检索和相关联想。

(3) 个性化智能兴趣代理

本系统的个性化智能兴趣代理是运行在系统的查询接口模块中,当学生提出查询请求,信息检索模块通过检索将概念进行扩展后,智能代理再根据学生的个人兴趣模型提取学生感兴趣的信息,并将检索结果呈现给学生。由于该搜索引擎是在网络教育资源环境下,面向网络学习者,而对于这些学生来说,其学习行为基本上是在本学科之内,其兴趣范围较Internet用户稳定得多,因此可以在语义网络上形象地构建学生个人兴趣网络。

基于上述思想,我们提出了一种全新的个性化智能兴趣代理系统,即在概念语义网络的

基础上, 通过学生对检索结果的反馈信息, 逐渐建立各概念节点的横向联系, 对学生的学科子树上的各节点及节点之间关系进行等级计算, 得到学生兴趣点的概念和关联等级排序。这样不仅可以得到学生感兴趣的关键词, 还能得到学生所感兴趣的一组相互有关联的兴趣词, 以此来确定学生的兴趣趋向。

4. 系统特点

本系统具有以下主要特点:

(1) 自动过滤不相关文档, 提高检索精度和效率。

(2) 采用“概念树”结构, 实现关键词在概念方面的检索, 将目前基于词的检索提高到概念层次。

(3) 运用语义网络构建“概念网”, 实现了概念层次上的同义扩展检索、概念拓展检索和相关联想, 提高了系统的查全率。

(4) 通过个性化兴趣代理建立学生的个性化兴趣模型, 过滤出学生所需要的信息资源, 提高了系统的查准率。

(5) 能够快速自适应用户兴趣和环境的变化。

结束语

本文讨论的个性化智能搜索引擎虽然是运用在特殊领域中, 但其研究内容属于目前智能信息检索领域的重要课题, 具有很强的理论意义和实际意义。在本文中, 我们对构建个性化、智能化的搜索引擎提出了一些新观点和方法, 如构建概念网和个性化智能兴趣代理等。但系统中还是有很多有待进一步研究和解决的问题, 如: 系统虽然能通过概念网对查询领域词进行概念扩展, 方法简单有效且易于理解, 但也只是在查询句分词、提取领域词的基础上扩展, 始终还不能完全理解学生的查询请求, 尤其不能理解学生的查询语句的语法结构, 应该在句法的理解上进行深入研究, 最好是从语义上进行理解。进一步开展以领域本体为知识背景构建专业化的智能搜索引擎, 将是我们今后研究工作的方向。

[参考文献]

- [1] 汪晓岩, 胡庆生, 李斌等. 面向Internet的个性化智能信息检索[J]. 计算机研究与发展, 1999, (9): 1040-1046.
- [2] 李伟超, 牛改芳. 智能代理技术分析及应用[J]. 情报杂志, 2003, (6): 29-33.
- [3] 董明杰, 张萍. 基于个性化信息服务的网络信息资源管理[J]. 图书情报工作, 2002, (6)
- [4] 饶增阳. 网络环境下的个性化信息服务[J]. 情报探索, 2004, (3): 3-4.
- [5] 张宏斌, 朱明富, 陈德军. 智能化搜索引擎技术的研究进展[J]. 信息与控制, 2003, (6)
- [6] 黄曾阳. HNC(概念层次网络)理论[M]. 北京: 清华大学出版社, 1998.

作者简介：吴丽华，硕士；罗云锋，教授，博士生导师。海南师范大学计算机系
(571158)。

(发表于《中国远程教育》，2007年第7期)

[【资料】](#) [【短消息】](#) [【订阅】](#) [【收藏】](#) [【我要发布】](#) [【评论】](#)



Copyright © 2007 本网站版权归：徐州师范大学|教育技术学科网,未经同意严禁转载、镜像。

[清除 Cookies](#)|[联系我们](#) | [关于我们](#)

地址：徐州师范大学信息传播学院 (221009)