

基于领域本体库的电子政务智能浏览系统

葛新¹, 董朝阳²

(1. 西安未来国际信息股份有限公司, 陕西 西安 710075;

2. 西安建筑科技大学机电工程学院, 陕西 西安 710055)

摘要: 为了使用户能够快速、智能地利用电子政务信息资源, 文中对不同来源、不同结构的数据进行整合和处理, 提供面向业务领域主题查询、检索的智能浏览系统, 给出了电子政务智能浏览系统体系结构, 对系统的主要功能如数据发掘、数据存储、数据挖掘分析、智能浏览服务器、智能浏览客户端等进行了分析; 对系统实现的关键技术即基于领域本体库的电子政务智能浏览进行了深入研究, 讨论了领域本体实例库的建立方法, 分析了智能浏览的工作流程, 并对领域本体的构建、推理机制、智能浏览等关键活动进行了研究。

关键词: 领域本体库; 电子政务; 智能浏览

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2012)08-0131-04

Intelligent Browsing System for E-Government Based on Domain Ontology Base

GE Xin¹, DONG Zhao-yang²

(1. Xi'an Future International Information Co. Ltd, Xi'an 710075, China;

2. College of Mechanical and Electronic Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: In order to use the information of e-government rapidly and smartly for users, the data of various sources and structure are integrated and processed. The intelligent browsing system is provided, which can be used for querying and retrieving of domain topics. The architecture of intelligent browsing for e-government is proposed. The main parts of the system are analyzed, such as data mining, data storing, data analyzing, intelligent browsing server, as well as intelligent browsing client. The key technology of the system, namely the intelligent browsing based on ontology is discussed. The constructing method for ontology instance base is proposed. Finally, the work flow of intelligent browsing is analyzed, and the key activities are studied, such as construction of domain ontology, deducing mechanism, as well as intelligent browsing.

Key words: domain ontology base; e-government; intelligent browsing

0 引言

陕西省电子政务建设经过多年的努力, 已经取得了较大的成就, 建立了大量应用系统并且建立了众多集中统一的数据库, 例如陕西省企业基础数据库、陕西省法人数据库、陕西省人口基础数据库、陕西省宏观经济数据库等, 电子政务建设发展的高级阶段就是对众多数据库中信息资源进行共享和整合, 以期更好地利用电子政务中的信息资源, 通过面向业务领域主题的查询、检索的智能浏览系统平台, 为用户提供快速、智

能地利用资源; 同时使得信息服务资源具备计算机互操作和系统的智能识别。

1 电子政务智能浏览系统体系结构

电子政务智能浏览系统主要包括数据源、数据发掘、数据存储、数据挖掘分析、智能浏览服务、智能浏览器、安全支撑服务等组成部分, 如图1所示。

1.1 数据发掘

1) 数据抽取。

数据抽取的作用是将不同来源、不同结构、不同数据介质的数据通过抽取机制汇集成有限的几种数据格式, 便于数据预处理模块对数据进行预处理分析。

结构化数据发掘针对数据库进行抽取, 该模块直接读取多级数据交换系统汇集好的数据库作为结构化数据库。空间数据发掘针对电子地图数据进行抽取,

收稿日期: 2011-12-30; 修回日期: 2012-04-03

基金项目: 陕西省工业攻关项目(2010K09-05)

作者简介: 葛新(1972-), 男, 技术总监, 主要研究方向为软件技术及电子政务; 董朝阳, 博士, 副教授, 主要研究方向为软件技术及智能优化算法。

空间数据包括地理信息和与地理相关的业务统计数据,该模块直接读取多元数据交换系统中的空间数据信息。非结构化数据主要通过读取公文、表格等非结构化的文件资料,从文件资料中抽取所关注的元数据进行建库的工作。

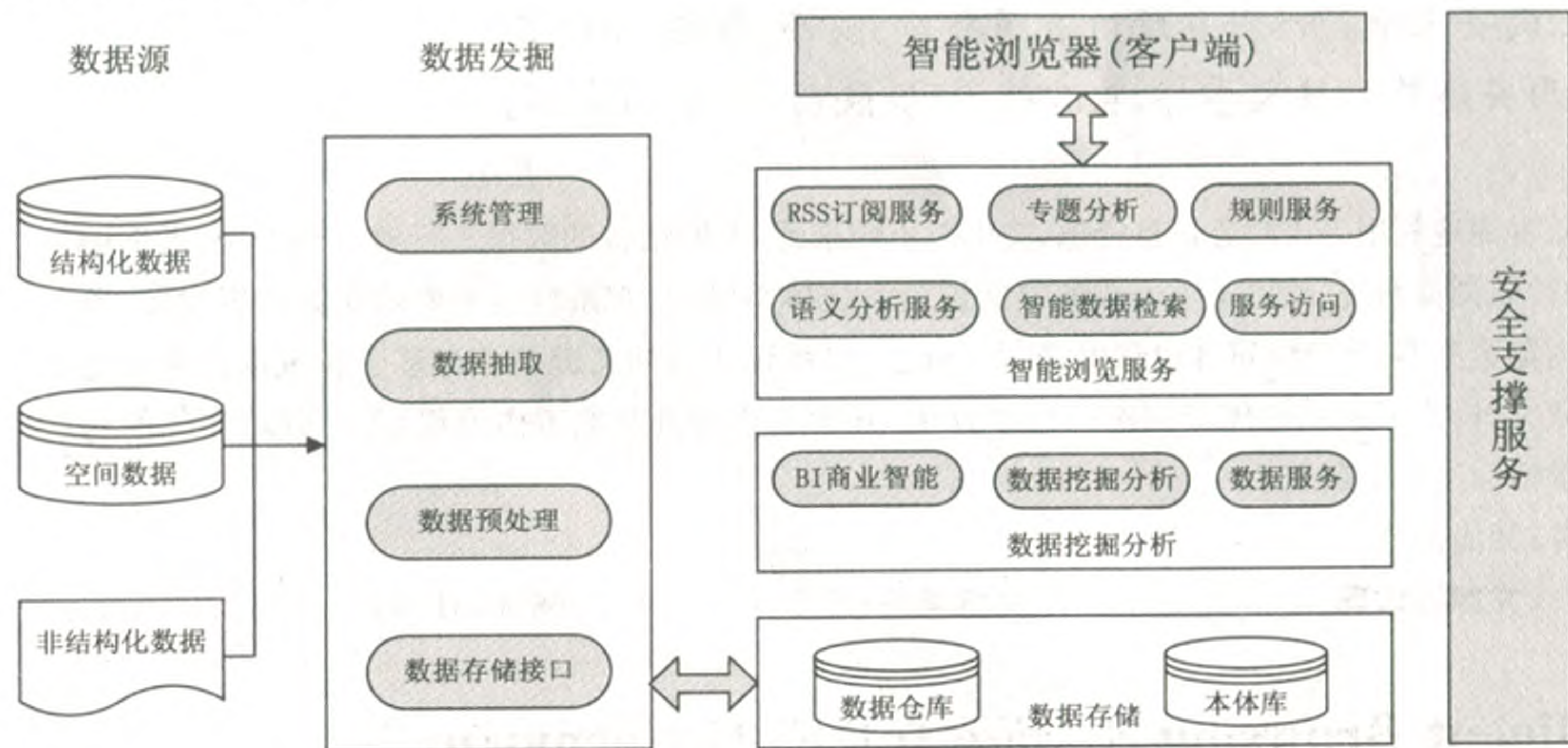


图1 电子政务智能浏览系统体系结构

2) 数据预处理。

根据预先制定的预处理规则对数据抽取模块汇集的数据进行预处理,使其满足数据仓库的数据结构要求,预处理完成的数据通过数据存储模块写入数据仓库。这是一个基本而重要的程序,其中包括与各个数据源及用户共同制定审核数据源的数据定义、数据的标准化及格式化、在标准化过程中需制定数据转换的商业规则、以及数据仓库的模型映射。

1.2 数据存储

数据存储区是为了保证数据移动的顺利进行而开设的永久性数据存储空间,需要进入数据仓库的各个业务系统的数据首先直接快速传输到数据转储区,再从数据存储区经过清洗、转换、映像等复杂的数据移动处理转移到目标数据仓库中。数据存储区保存数据仓库中最多最详细的业务数据,该层数据属于未经汇总的数据,但数据的组织方式可能已经完全不同于业务系统。根据业务需求的不同,基础数据仓库的组织形式以三范式模型为主,在有的系统中也可能采用星型或雪花型,由于基础数据仓库数据是对原始业务数据的原形再现,所以数据量会非常庞大,根据业务需求的不同,数据保留时间在一年到多年不等。

数据存储提供标准接口供数据预处理模块读写数据仓库的数据。该模块提供了针对基于关系数据库构建的数据仓库的数据访问接口、针对海量分布式文件系统提供文件访问接口、以及针对本体操作的访问接

口。

1.3 智能浏览服务

智能浏览服务作为智能浏览器的后台支持服务系统,用于在服务器端接收浏览器服务请求,收集并组织数据以智能浏览器能识别的报文格式发送给智能浏览

器供浏览器展示信息。智能浏览服务模块与系统其它模块的关系如图2所示。

智能浏览服务的主要功能包括:

1) 用户配置信息。

用户可以注册和管理个人信息,并保存用户对指定业务应用的存储属性。

2) RSS 订阅服务。

用户可以在线添加 RSS 订阅源,智能浏览服务汇集用户订阅的 RSS 信息源,动态更新 RSS 信息,并向用户主动通知订阅现状。

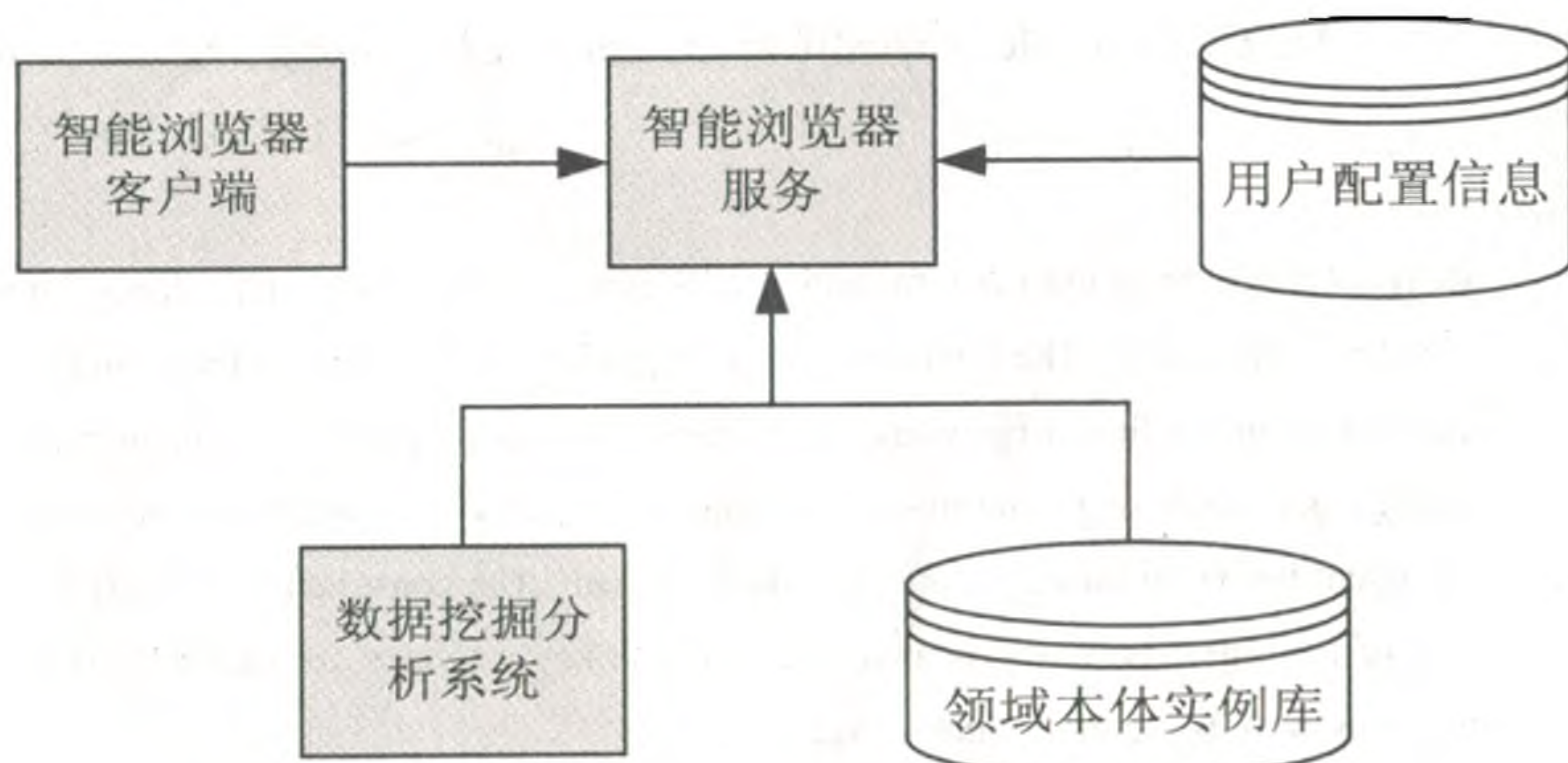


图2 智能浏览器与服务端程序架构

3) 专题分析。

根据数据仓库和本体库的数据内容,提供相应的专题分析查询。例如,人员专题分析:根据人口库和企业法人库等数据库查询分析该人员的综合个人信息和社会关系信息;企业关系分析:根据企业数据库查询分析企业运营数据等。

4) 规则服务。

可以预定用户访问规则,根据预定的时间,地理规则控制用户访问内容。例如,设定用户访问规则,根据用户权限过滤访问内容;设定工作时间,用户只能在工作时间访问涉密专题分析;设定 IP 地址范围绑定,用户只能在指定 IP 网段访问涉密专题分析;综合用户身份,时间和空间的属性设定综合访问规则。

5) 智能数据检索。

提供数据搜索功能,根据数据仓库、数据挖掘、文

件数据、本体库属性进行搜索并向用户提供数据搜索结果。用户可订制智能搜索内容,当数据仓库中有新数据符合定制智能检索条件时,向用户主动推送数据。

1.4 数据挖掘分析

数据挖掘是从大量数据集合中有效地自动发现隐含的、先前未知的、对决策有潜在价值的信息,数据挖掘从起源本质上就与统计学、模式识别、数据库、人工智能、高性能与并行计算等学科有着密不可分的关系^[1,2]。数据挖掘基于数据仓库,即业务数据经过提取、滤波、清除和聚集后,利用“叠加”方法增加有关方面字段,重新组织成为单独存储的数据库;使用复杂的统计分析和建模技术来揭示数据库中隐藏的模式与关系,从数据的分析入手帮助决策,并从数据中寻找有价值的规律的技术;除要求提供一般的数据分析处理要求以外,数据挖掘还包括关联/序贯分析、聚类分析、分类分析、回归分析等主要功能。在本系统中,数据挖掘分析采用第三方 BI 系统实现 BI 智能、数据挖掘分析、数据服务功能。

1.5 智能浏览客户端

由于单纯的 web 浏览器受功能和安全性的限制只能实现有限的功能,为了能更好利用用户桌面的性能,需要开发更有效和更智能化的智能浏览器客户端。开发智能浏览器客户端采用 Adobe AIR 技术。Adobe AIR 是一个跨操作系统运行时,它使开发人员能够将 HTML、Ajax、Adobe Flash 和 Flex 技术结合到一起以在桌面上部署富 Internet 应用程序(RIA)。智能浏览器客户端具有以下特点:

1) 使用本机文档处理程序打开文档。

使用与文件关联的本机应用程序打开 PDF、PSD、DOC、PPT 和 MP3 等常见文件格式的文档。

2) 安全性改进。

设置应用的可执行权限。充分利用加密的 TLS 套接字和对 MAC 地址的访问权。

3) 本地嵌入式数据库。

使用同步和异步 API 管理本地数据库中的结构化数据。

4) 增强的 HTML 安全性。

可为在桌面上的 HTML 应用程序提供更高的安全性。这些改进减少了常见 web 漏洞(如跨站脚本和代码注入)对应用程序构成的潜在威胁。

5) 网络内容沙箱。

为所有网络内容提供一个沙箱,它类似于浏览器沙箱,用于限制网络内容对运行时完整功能的访问。

6) 网络检测。

能够检测到它们何时连接到网络。脱机时,数据输入可以存储在本地,然后在重新连接时与服务器同步。

2 建立本体实例库

智能浏览的一个重要环节就是通过语义相似度^[3-5]检索领域本体实例库,本体实例库的建立是决定智能浏览功能实现的关键,本体实例库建立的过程如图3所示:

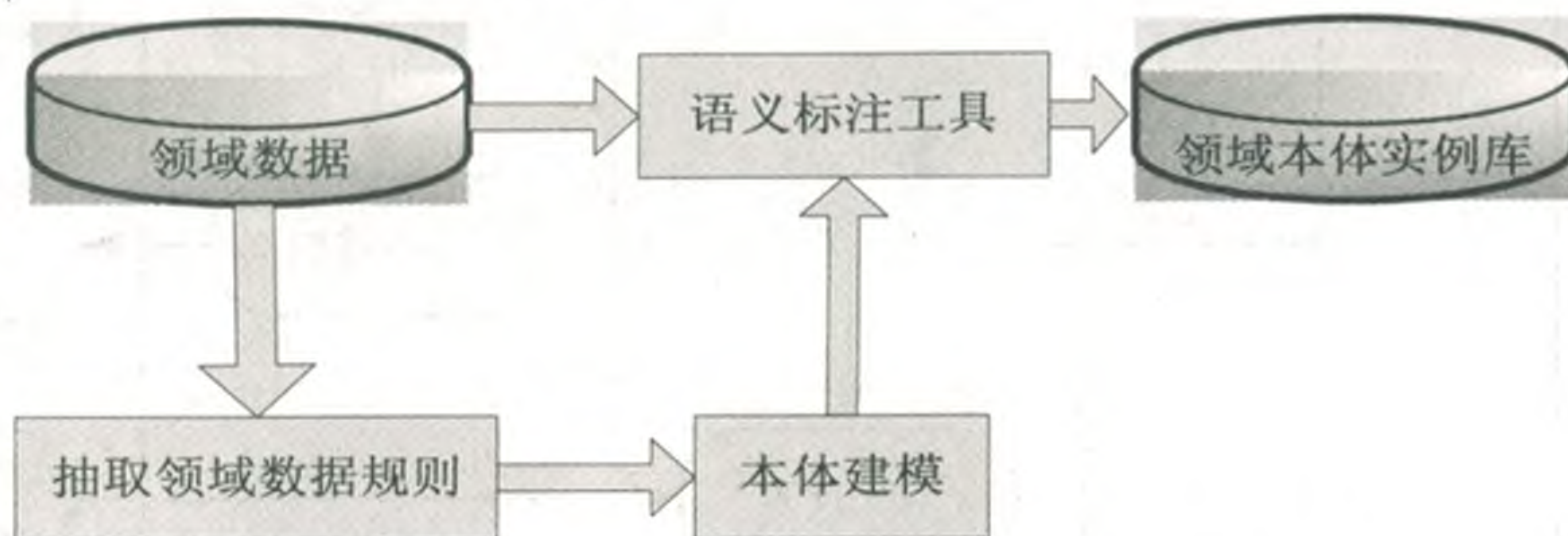


图3 领域本体实例库建立过程

通过领域数据可以抽取出领域数据规则,并通过 RDFS 建模从而建立起领域本体模型,结合领域本体模型并通过自动化语义标注工具,就可以对大量领域数据进行标注,每一个标注就是一个本体实例,所有本体实例共同构成了领域本体实例库,以便于用户的检索。

3 智能浏览系统工作过程

智能浏览系统主要的活动如图4所示,下面对其中的关键问题予以讨论。

1) 领域本体的构建。

本体的构建是一个不断修改和精炼的过程^[6-9],本系统从数据仓库中发掘本体对象,依据当前可获得数据源的类型,目前系统已经实现了三种本体的发掘构建工作:“人员”本体、“企业”本体、以及“公文”本体等。

2) 推理机制。

在智能数据浏览系统中,用户可以输入所要搜索要求,再由语义推理机利用概念相似度求出查询请求向量和本体实例库中各本体实例向量的语义相似度,最后根据语义相似度的大小将检索结果筛选、排序返回给用户。

语义相似度计算方法包括基于点的方法、基于边的方法、基于概念树的方法^[10-12]。基于点的相似度计算方法通过概念及其实例对象出现概率或权重来量化;基于边的相似度计算方法主要考虑的是概念之间的距离,实际上就是计算这两个概念在多大程度上共享一个语义信息内容;基于概念树的方法不仅计算两棵概念树中两点的语义相似度,还要计算两棵概念树

之间的语义相似度。在本系统中,语义相似度采用基于概念树的方法进行计算。

discovered rules from data mining by DEA[J]. Expert Systems with Applications, 2009, 36(4): 8503-8508.

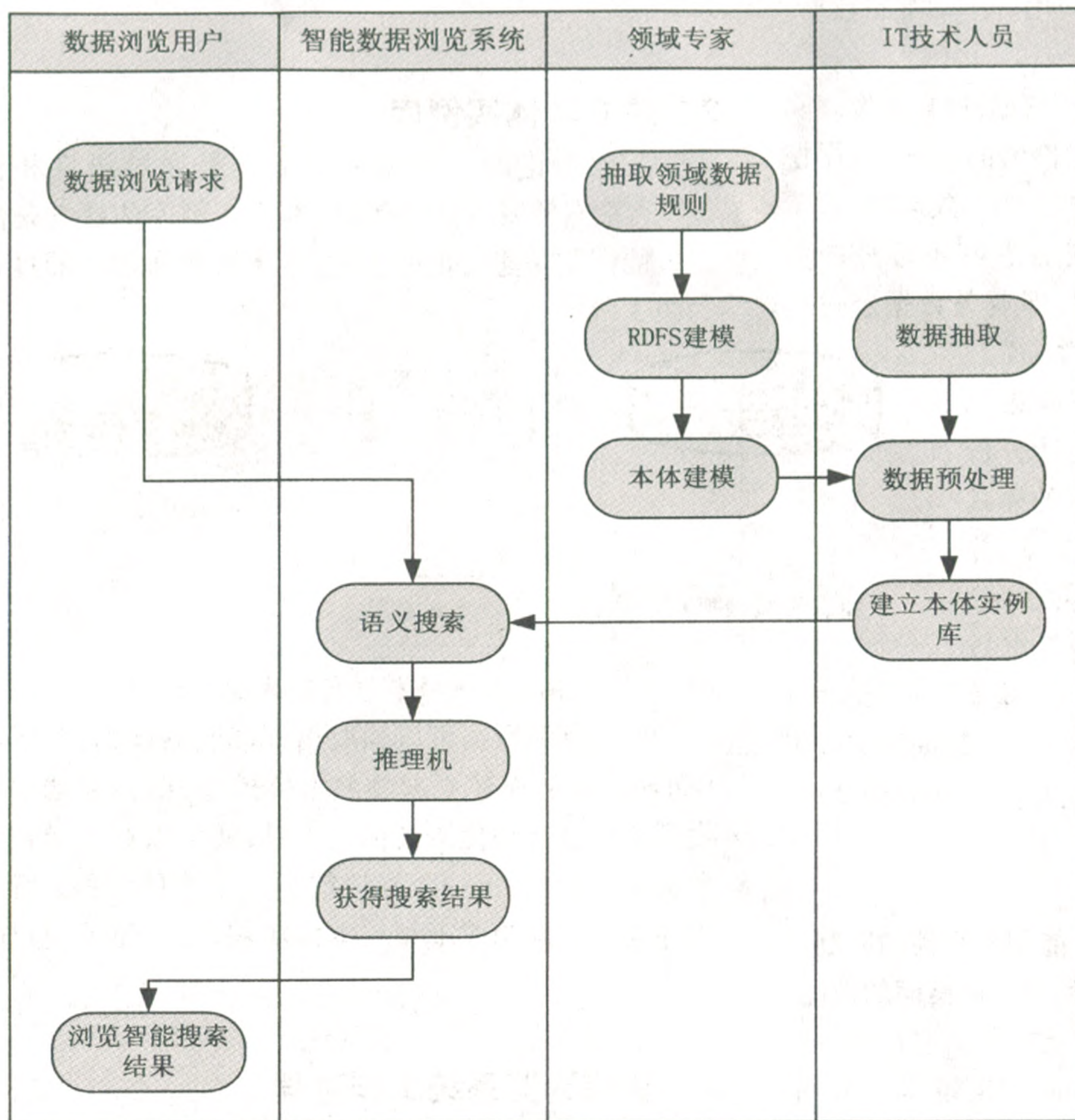


图 4 智能浏览系统主要活动

3) 智能浏览。

智能浏览面向终端用户,提供诸如 RSS 订阅服务、专题分析、规则服务、语义分析服务、智能数据检索、服务访问等服务,其中智能数据检索是关键而复杂的服务。

4 结束语

文中研究成果已经用于陕西省电子政务智能浏览系统的建设中,该系统是陕西省电子政务支撑中心建设项目的重要组成部分,体现文中思想的电子政务智能浏览系统的成功实施必将为提升陕西省电子政务整体建设水平和业务应用的易用性起到重要作用。

参考文献:

[1] Toloo M, Sohrabi B, Nalchigar S. A new method for ranking

[2] Ning Xiaomin, Jin Hai, Wu Hao. RSS: A framework enabling ranked search on the semantic web[J]. Information Processing and Management, 2008, 44(2): 893-909.

[3] Beydoun G. Formal concept analysis for an e-learning semantic web[J]. Expert Systems with Applications, 2009, 36(8): 10952-10961.

[4] Conesa J, Storey V C, Sugumar V. Improving web-query processing through semantic knowledge[J]. Data & Knowledge Engineering, 2008, 66(1): 18-34.

[5] Du T C, Li Feng, King I. Managing knowledge on the Web-extracting ontology from HTML Web[J]. Decision Support Systems, 2009, 47(4): 319-331.

[6] 柳巧玲, 米天胜. 基于语义网的智能信息检索研究[J]. 科技管理研究, 2008(8): 160-162.

[7] 陈布伟, 李冠宇, 张俊, 等. 基于语义网规则语言的推理机制框架设计[J]. 计算机工程与设计, 2010, 31(4): 847-853.

[8] 赵欢, 李仁发, 王家琴. 综合多层信息的本体概念相似度计算方法的研究[J]. 通信学报, 2009, 30(6): 135-141.

[9] 黄敏, 赖茂生. 语义检索研究综述[J]. 图书情报工作, 2008, 52(6): 63-66.

[10] 陶望胜, 陶先平, 吕建. 基于规则与相似度的语义 web 服务匹配[J]. 南京大学学报(自然科学), 2010, 46(2): 159-168.

[11] 吴江宁, 杨光飞. 基于本体的项目和领域专家匹配原型系统[J]. 计算机应用研究, 2009, 26(10): 3787-3790.

[12] 张亮, 屈振新, 丁蕊, 等. 一种基于加权领域本体的语义检索方法[J]. 计算机科学, 2010, 37(7): 165-168.

Ad Hoc 网络中基于能量的 QoS 路由协议的研究

作者:
作者单位:
刊名:
英文名:
年, 卷(期):

章晓, 束永安
安徽大学计算机科学与技术学院, 安徽合肥230039
计算机技术与发展
Computer Technology and Development
2012 (8)

参考文献(12条)

- 1.Toloo M;Sohrabji B;Nalchigar S A new method for ranking discovered rules from data mining by DEA[外文期刊] 2009 (04)
- 2.Ning Xiaomin;Jin Hai;Wu Hao RSS:A framework enabling ranked search on the semantic web 2008 (02)
- 3.Beydoun G Formal concept analysis for an e-learning senantic web 2009 (08)
- 4.Comesa J;Storey V C;Sugumaran V Improving webquery processing through semantic knowledge 2008 (01)
- 5.Du T C;Li Feng;King I Managing knowledge on the Web-extracting ontology from HTML Web 2009 (04)
- 6.柳巧玲;宋天胜 基于语义网的智能信息检索研究[期刊论文]•科技管理研究 2008 (08)
- 7.陈布伟;李冠宇;张俊 基于语义网规则语言的推理机制框架设计[期刊论文]•计算机工程与设计 2010 (04)
- 8.赵欢;李仁发;王家琴 综合多层信息的本体概念相似度计算方法的研究[期刊论文]•通信学报 2009 (06)
- 9.黄敏;赖茂生 语义检索研究综述[期刊论文]•图书情报工作 2008 (06)
- 10.陶望祖;陶光平;吕建 基于规则与相似度的语义web服务匹配[期刊论文]•南京大学学报(自然科学版) 2010 (02)
- 11.吴江宁;杨光飞 基于本体的项目 and 领域专家匹配原型系统[期刊论文]•计算机应用研究 2009 (10)
- 12.张亮;胡振新;丁蕊 一种基于加权领域本体的语义检索方法[期刊论文]•计算机科学 2010 (07)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjfr201208034.aspx