

Web 数据仓库及其在办公自动化系统中的应用

王昭义, 刘 斌, 蔡瑞英

(南京工业大学 信息科学与工程学院, 江苏 南京 210009)

摘 要:随着办公自动化系统应用的日益深入,产生了大量的历史数据,同时,在办公自动化系统中,也需要利用历史数据和环境数据进行办公自动化的辅助决策。因此可以利用 Web 数据仓库技术,将办公自动化系统中的海量数据重组,构成可用、可控、可扩展的数据组织,以满足企业各级主管和办公、业务人员分析决策的信息需求。文中介绍了 Web 数据仓库在办公自动化系统中的应用。在分析 Web 数据仓库以及办公自动化的概念及特点的基础上,结合办公自动化系统的数据形式和决策需求,描述了 Web 数据仓库技术与办公自动化系统结合的框架结构,并以一个子系统为例说明 Web 数据仓库技术在办公自动化系统中的具体实现方法。

关键词:Web 数据仓库;办公自动化;数据挖掘

中图分类号:TP311.13

文献标识码:A

文章编号:1005-3751(2006)02-0017-03

Web Data Warehouse and Its Application in OA

WANG Zhao-yi, LIU Bin, CAI Rui-ying

(Coll. of Info. Sci. and Eng., Nanjing Univ. of Tech., Nanjing 210009, China)

Abstract: As the OA system application increasingly widely used, there are large history date. At the same time, OA system needs history date and surrounding date to assistant decision. In order to meet analysis decision information demands of director and office man, it could use Web data-warehouse reform great date to form useable patulous manipulative date. This paper introduced the application of Web data-warehouse in OA system. On the base of analyzing the concepts, characters of Web data-warehouse and OA, considering the data form and DSS demand of OA, described the architecture of combination of Web data-warehouse and OA. Also gave the realization of a subsystem to show how Web data-warehouse was used in OA.

Key words: Web data-warehouse; OA; data mining

0 引 言

数据仓库是一个面向主题的、集成的、不可更新的、随时间不断变化的数据集。当前,基于 Web 的网络技术正在改变人们的行为方式和工作模式,Web 数据仓库也应运而生,它是数据仓库与网络技术相结合的产物,它的出现为数据仓库的应用和发展开辟了新的天地。办公自动化系统将计算机技术、通信技术、系统科学、行为科学应用于传统的数据处理技术难以处理的、结构不明确的业务^[1],它能够充分有效地利用信息资源,改善管理水平,提高生产、办公质量和效率,增强企业竞争力。文中主要介绍 Web 数据仓库在办公自动化中的应用。

1 Web 数据仓库

Web 数据仓库是数据仓库和 Web 技术相结合的产物,数据仓库和 Web 技术为 Web 数据仓库提供了概念体

系和技术基础^[2]。数据仓库为查询和决策分析提供集成化信息仓库^[3],是企业界的诊断系统。它对 Web 数据仓库的贡献在于:首先,它为 Web 数据仓库提供完善的技术基础,几乎所有运行在数据仓库中的技术仍可应用于 Web 数据仓库,这些技术包括查询、OLAP、代理、统计分析、数据挖掘、绘图/图像系统等;其次,为 Web 数据仓库提供信息的管理规则和管理目标,数据仓库通过信息资源的管理使企业的分析工作变得简单,无论需要什么信息以及何时需要信息,数据仓库都能提供,将这一特性应用于 Web 数据仓库,使其能够有效地管理各种不同的信息资源。但是,Web 数据仓库也有自身的特点,即原来运行在数据仓库中的应用和信息传输在 Web 数据仓库中是通过 Web 技术实现的。

基于分布式结构的 Web 技术不仅为用户传递信息带来新的方式,而且使用户能够操纵不同的信息和对象,且信息和对象更易获取和传递,不需要太多的技术就可以进行查询、生成记录和智能分析。Web 技术使决策者在信息共享和协同分析方面得到突破,并且改变了支持决策的数据表达和数据分析方式。Web 数据仓库正是通过采用新的网络技术,将 Web 技术和企业智能系统相结合,适用

收稿日期:2005-05-27

作者简介:王昭义(1973—),女,浙江绍兴人,硕士研究生,研究方向为进行数据仓库和数据挖掘的研究与应用;蔡瑞英,教授,研究方向为人工智能与网络服务。

于多种工作环境。另外,由于引进 Web 技术,使得 Web 数据仓库对时间敏感性和地域敏感性要求较低。

Web 数据仓库是环境以及技术变更之后的数据仓库发展的结果,或者说,是数据仓库和 Web 技术的结合所产生了更新的、更具倾向性的、更便于协作管理的一种辅助决策技术。

Web 数据仓库的体系结构如图 1 所示。

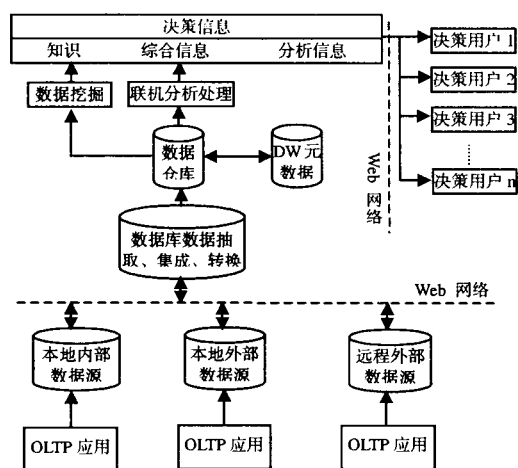


图 1 Web 数据仓库的体系结构

2 办公自动化

随着信息技术的发展和国内信息技术应用水平的提高,全面实施办公自动化已经是提高政府机关和企事业单位办公效率以及实行规范化管理的必然趋势。首先,OA 可以减少重复劳动,提高工作效率;其次,OA 可以实现工作程序流程化,内部文档流转无纸化,加快办公文档的周转和处理速度,降低文档传递中的失误率,使各项工作基本达到规范化、标准化;再次,OA 可以有效提高信息的综合利用能力,增强综合分析能力;最后,OA 可以通过对文档流转中的监控,能够及时发现和掌握工作环节中的瓶颈,为科学管理提供可靠的依据。

从基于 Web 技术和面向知识管理的第三代办公自动化系统的特征来看,将 Web 技术和知识管理应用到办公自动化上,不仅把各种资源实体化和量化,实现政府机关、企事业单位内部的办公电子化和网络化,而且各部门可以以知识管理为基础,让全体工作人员和各部门之间实现信息共享和实时通讯,从而建立一个以反映和满足工作需求为导向的电子化办公平台,为政府机关及企事业单位建立了一个高效、实用、广泛、快捷的知识管理服务平台。

3 Web 数据仓库在 OA 中的应用

与其它应用系统一样,办公自动化系统也包括信息的采集、加工、传输、保存 4 个环节。随着计算机应用技术的进步和经济建设的发展,信息处理技术已从单纯的数据处理发展到信息的管理和加工,当前,以 Web 作为信息传输平台、以联机分析处理和数据挖掘等工具为手段的 Web

数据仓库的辅助决策功能为办公自动化注入了新的活力。

将 Web 数据仓库应用到办公自动化系统中,使办公自动化更加智能化^[4]。在 Web 数据仓库的 DSS 中,数据仓库是整个 DSS 的基础,它从底层的事务应用产生的数据(数据源)中抽取数据,由于这些数据主要来自分布式环境,它们分布在不同的存储介质、数据库以及不同网络拓扑结构中,为不同的业务部门服务。从结构上看,这些业务部门的数据是相对独立的,这是不利于企业决策者进行全面分析和查询的,因此必须对其进行综合、集成与转换(抽取、集成、转换...)。但是如果仍然使用传统的数据共享、正规化等思路来组织和管理这些数据,很难将这些数据转化为集中统一的、随时可用的信息。基于 Web 数据仓库的决策支持系统运用反正规化思路,引入辅助决策把数据集成为一个统一的整体,提供面向全局的数据视图^[5]。

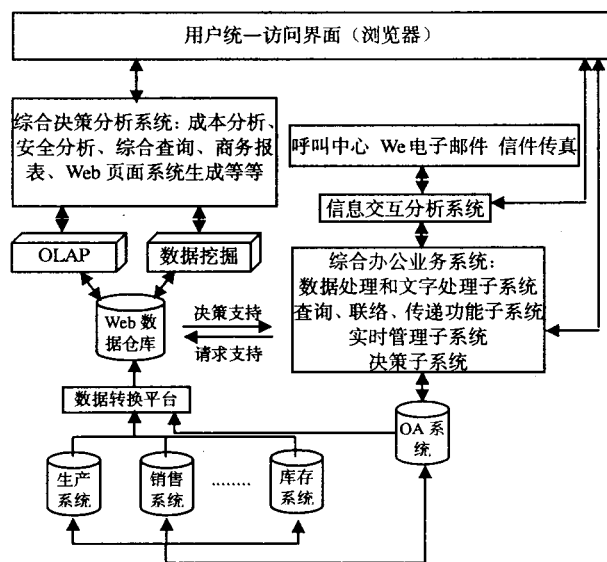


图 2 Web 数据仓库在办公自动化系统中实现

图 2 为 Web 数据仓库系统应用于 OA 系统的体系结构图,从图中可以看出,整个系统由 3 大功能模块构成:即 OA 系统、综合决策分析系统以及互动信息分析系统。OA 系统主要承担底层事务管理工作,根据各部门职能分工不同,OA 系统可划分为不同的子系统,主要包括数据处理和文字处理子系统、查询联络等传递功能子系统、实时管理子系统(如图 3 所示)、判断决策树子系统等。一般来说,在不同子系统的数据具有不同形式和特征:与文字处理相关的数据,包括在生成文件、信息检索、计算等过程中所处理和产生的数据;与传递功能相关的数据,包括电话、会议、会谈、文件等的信息分发或传递等过程中所处理和产生的数据;与实时管理相关的数据,包括确定安排会议或会谈的时间、程序、地点等过程中所处理和产生的数据;与决策相关的数据,包括文件的报审、批准,问题的讨论、确定等过程中所处理和产生的数据。这些数据的获取、处理和加工都需要 Web 数据仓库提供相应的决策支持。

综合决策分析系统以 Web 数据仓库为核心提供了丰富的辅助决策手段,为 OA 系统提供决策支持。Web 获得决策信息的主要手段为联机分析处理(OLAP)和数据挖掘(DM)。OLAP 提供切片(Slice)、旋转(Pivot)等在线分析方法,包括对生产、销售以及库存系统中报表上报和汇总情况、资金流动情况、办公用品的需求和使用情况、文件发放效率以及反馈信息、出差以及差旅费的使用情况等信息提供多角度实时查询及分析等功能,并获得高度归纳的分析结果,为用户提供决策支持,同时提供预警功能,防止资金非法流动、耗材库存不足以及恶意攻击等。DM 可以挖掘 OA 中隐藏、非直观的信息,预测未来的趋势,并可以直接用于指导 OLAP。

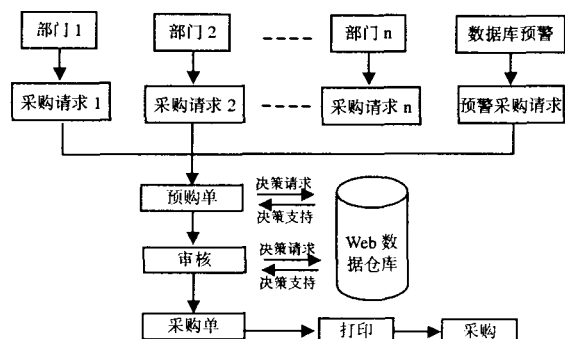


图3 实时管理子系统—办公用品采购

DM 对客户数据的分析过程分为数据准备、挖掘操作、结果表述 3 个阶段,事实上,这 3 个过程是一个不断递归执行的过程,即当分析决策人员对分析结果不满意时,就可以反复执行这 3 个过程,直到得到最有价值的信息,做出归纳性的推理,从中挖掘潜在的模式,预测 OA 系统中最佳工作模式和流程,为决策者做出正确判断和决策提供强有力的支持,实现 OA 系统的效率最大化和资源最优配置。

交互信息分析系统主要对客户行为进行 Web 数据挖掘,获得用户的行为习惯,并检测恶意用户。这类 Web 数据挖掘包括结构挖掘、应用挖掘和内容挖掘 3 种行为。结构挖掘用于提取网络的拓扑信息——网页之间的链接信息;应用挖掘用于提取关于用户如何运用浏览器浏览和使用这些链接的信息;内容挖掘是用来提取文字、图片或其它组成网页内容成分的信息^[6]。

(上接第 16 页)

频交互功能,增加对综合电信业务办理的支持。提出了一种基于服务时间和业务熟悉程度相结合的多队列机制来实现坐席分配。交互式自助服务平台的实现对于电信运营商树立自身形象,提高服务质量,增加经济效益,减少投资成本,增强竞争力具有很重要的意义。

参考文献:

- [1] 于海燕,张霞.基于 TUXEDO 中间件的邮政电子化支局生产作业系统的设计与实现[J].计算机系统应用,2003

Web 数据仓库强大的决策分析功能是 Web 数据仓库在办公自动化综合决策支持系统中实现的灵魂,它使办公自动化系统具有一定的智能特性,能优化和量化 OA 中的各种资源。此外,信息交互系统是决策者或决策部门与办公自动化系统打交道的界面,它负责接收用户发出的各种指令(比如,对确定的主题进行分析、对比、预测等决策处理;对决策辅助数据进行各种查询等等),根据这些指令调用不同的子系统,并获得处理结果,最后将结果输出给用户。

4 结束语

基于 Web 的办公自动化系统提供了一个可视化的、高效的、实用的企业办公平台,它可以在企业现有信息子系统的基础上,利用 Web 技术便捷地为企业构架面向管理决策层的智能办公自动化系统,以提高企业的办公效率,降低办公成本。但由于现有数据仓库中的数据挖掘技术必须建立在结构良好的数据基础之上^[7],而 Web 上存在许多半结构化和非结构化数据,因此,Web 数据仓库还有待进一步发展。另一方面,由于数据量的限制,将 Web 数据仓库应用于办公自动化还处于起步阶段,随着数据的积累,这一领域将会受到越来越多人的重视。

参考文献:

- [1] 肖春芸.基于 Web 的数据仓库[J].计算机与现代化,2001(2):136-139.
- [2] Xin T. Web warehousing: Web technology meets data warehousing[J]. Technology in Society, 2003, 25(1): 133-139.
- [3] 李秀,廖璘,刘文煌.基于 Web 的数据仓库系统的研究[J].计算机工程,2001,27(11):44-46.
- [4] 张庆,董红斌.基于 Web 的数据仓库技术及其在商业银行中的应用[J].计算机应用研究,2002(5):88-91.
- [5] 陈文伟.决策支持系统教程[M].北京:清华大学出版社,2004.
- [6] Linoff G S, Berry M J A. Web 数据挖掘:将客户数据转化为客户价值[M].沈钧毅,宋擒豹,等译.北京:电子工业出版社,2004.
- [7] 朱红,王兆锐,由颖.基于 Web 的数据挖掘模型[J].沈阳工业大学学报,2002,24(1):61-63.

(7):53-56.

- [2] 罗会兰,谢小兵.多层分布式体系技术在电信触摸屏多媒体查询系统中的应用[J].微型机与应用,2004(1):36-38.
- [3] 张海林,赵锦蓉,庄永宙. Multimedia Conferencing System and Multicasting[J]. 清华大学学报,1999(2):35-39.
- [4] 郭亚,刘晓晴.基于 Web 技术的自助服务开发平台[J].中国金融电脑,2003,12:49-51.
- [5] Seok - Soo kim, Dae - Joon Hwang. CSCW based telemedicine on the Internet [A]. proceedings of the 32nd annual Hawaii international Conference on System Sciences [C]. Hawaii: IEEE, 1999. 5-8.