

Web Service 和信息检索技术在企业信息化中的应用

Application of Web Service and Information Research&Reference Technology for Enterprise Informatization

陆 远 张晓东 顾 名 (清华大学 软件学院 北京 100084)

摘要:从国内企业信息化的概念、发展现状以及存在的问题入手,针对数据分散,信息利用效率低下等问题,提出了基于 XML 网络服务和信息检索技术为基础的一个解决方案,并结合微软员工信息查询系统(EIS)的系统设计实例,介绍如何利用网络服务和信息检索技术解决实际问题。

关键词:信息检索 XML Web Service 企业信息化 EIS SOAP

1 企业信息化的现状和存在的问题

1.1 企业信息化简介

企业信息化从广义上理解是指:广泛利用电子信息技术,使企业产品设计、加工制造、过程控制、质量检测、生产管理和销售服务等活动的信息处理自动化。

企业信息化的核心任务主要有以下几个方面:

(1) 企业局域网及内联网建设与应用。

(2) 企业数据库建设与应用。

(3) 覆盖企业管理运营过程中各个方面的信息系统(或子系统)的建立与运行,包括 CAD/ CAM、MIS、DCS、CIMS、MRP、PDM、ERP、CRM 等。

(4) 企业电子商务系统的建立与运行。

(5) 企业人力资源、质保体系、服务体系、办公自动化文档管理系统(或子系统)的建立与运行。

(6) 企业资产经营、资本运作、子公司经营与管理等方面的信息建设与应用^[1]。

在信息时代,数据是企业的宝贵资源,企业信息化的目的也是能够高效的利用企业中的各种宝贵数据资源。企业数据资源来自内部数据和外部信息资料,包括诸多方面,如表 1 所示^[3]。

如何高效的利用和整合这些数据资源,成为当今许多企业信息化过程中遇到的一个难题。本文将重点放在信息处理和利用效率的问题上。

1.2 企业信息化中的信息处理存在的问题

在企业的信息化过程中,对于信息和数据资源的处理和利用,主要有以下的一些问题:

(1) 信息和数据的分散问题。虽然管理软件和数据库系统为企业的管理带来了很多方便,使系统内的信息非常丰富,但各系统之间独立有余,协同不足,各种信息无法在企业内部或外部有效传递或协同整合,结果形成了一个信息孤岛,使信息通信的效率低下。现实中,这种信息孤岛形成的问题,即存在于企业内部,也存在于不同的企业之间,例如:企业与企业之间的 MRP、ERP 等系统无法直接通信。

表 1

企业数据资源											
内部数据						外部数据					
产品图纸	技术文档	技术手册	生产营销	人财物	财务图表	人的经验	相关技术	行业信息	供应商	客户信息	政策法规

(2) 数据存储介质和信息系统之间存在鸿沟。仅针对电子数据和信息而言,目前,企业中的电子数据存储介质还是基于 Word、Excel、文本文件等这样的非规范化电子文档为主,这些文件蕴含了企业运作过程中最重要的文档和数据。要想有效的利用企业现存的信息系统整合这些信息难度很大,因为信息系统更倾向于依赖良好的数据库架构,对于数据输入源要求比较规范的格式。对于一些非规范数据更多的是采用剔除或者变通的处理方式,这大大增加了数据处理的代价,并降低了数据和信息处理效率和准确性。

(3) 用户界面操作体验以及可视化的问题。目前,随着企业信息化的展开,为了推广新的办公和信息系统,经常出现需要改变用户习惯情况。然而,改变用户的工作习惯代价是巨大的,这其中的培训成本,以及由此带来的员工抵触问题,都不容忽视。虽然一个出色的信息系统能够在日后的运作中去改变用户观念,但是这是一个代价极其昂贵,等待时间相当的过程。如果能够提供一种不改变用户使用习惯,而能够改进信息和数据处理效率的系统,将会大大提高用户的满意度和使用效率。

此外,由于大部分数据信息都是基于数字和文字表示方式的,这样的信息不利于用户快速的理解和检索需要的信息,使用难度大。特别是对于知识工作者而言,他们需要的不仅仅是大量杂乱的数据和信息,而是需要把它们快速转化成目标知识和技能。这就涉及到数据可视化的问题,这也是现在企业信息化的系统中还做得不够完善的地方。

1.3 企业信息化中对信息的高效利用

20 世纪 90 年代,因特网的迅速发展及普及为信息化和办公自动化开创了新局面。人们期望可以利用网络通信基础及先进的网络应用平台,最终实现“无纸”办公。

在实践中人们深切地认识到,即使企业实现了信息电子化和“无纸办公”,企业信息处理和利用的效率却并不能够让人满意,而只有充分完成了对信息和数据的高效率利用,才能够实现企业信息化的最终目的。

针对上面提到的三个问题,可以利用 XML 技术为基础的网络服务、可视化技术、Office 智能客户端和信息检索技术、数据挖掘等技术为企业信息化的进程提供便利,并真正为高效的信息和数据的处理提供良好的系统平台。

2 XML 网络服务和 Office 信息检索服务

2.1 XML 和网络服务概述

XML 即 Extensible Markup Language (可扩展标记语言)的缩写。XML 实际上是 Web 上表示结构化信息的一种标准文本格式。XML 语言有两大优势:一是自由,二是超越于格式之上^[2]。

XML Web Service 是在 Internet 上进行分布式计算的基本构造块。开放的标准以及对用户和应用程序之

间的通信和协作的关注产生了这样一种环境,在这种环境下,XML 网络服务成为应用程序集成的平台。应用程序是通过使用一个或多个不同来源的 XML 网络服务构造而成的,这些服务相互协同工作,而不管它们位于何处或者如何实现。

XML 网络服务通过标准的 Web 协议向 Web 用户提供有用的功能,多数情况下使用 SOAP (简单对象访问协议) 协议。SOAP 最引人注目的特征是它可以在许多不同的软件和硬件平台上实现。HTTP 的普及和 SOAP 的简单性使其成为 XML 网络服务的理想基础。

XML 网络服务可以非常详细地说明其接口,这使用户能够创建客户端应用程序与它们进行通信。这种说明通常包含在称为 Web 服务说明语言 (WSDL) 文档的 XML 文档中。WSDL (Web Services Description Language) 文件用于说明消息格式的表达法以 XML 架构标准为基础,这意味着它与编程语言无关,而且以标准为基础,因此适用于描述从不同平台、以不同编程语言访问的 XML 网络服务接口。

2.2 XML 网络服务在企业信息化中的应用

在许多企业中,都有多种系统和数据服务器并存的情况,如何在这些系统之间进行方便的通信是最大的难题。随着 XML 和网络服务应用的发展,只要它们能够分析 XML 并理解 SOAP 规范的规则,任何两个系统之间都能够进行交流。

通过 XML 网络服务的平台无关性,可以整合许多其他类型的企业资源,例如:ERP、CRM、DBMS 一起其他文件系统等,并提供一个统一的接口。这样企业的上层应用,可以针对单一的接口模式进行应用开发,这种对数据透明的结构,能大大提高企业上层应用开发和实施的效率和效果。通过网络服务整合了各种企业数据系统和信息资源,有效地消除了格式不一致,不同系统之间无法进行有效的信息沟通的问题。

如图 1,显示了在企业中,利用 XML 网络服务整合企业信息系统的通用体系结构。

该结构以 XML 网络服务为核心,整合了企业的内容服务 (包括 ERP、CRM、DB、Files、Internet 系统等),并为上层应用提供了一个统一的接口:XML Web Service Interface,这是一个开放的标准接口,上层应用程序可以通过这个接口访问各种需要的资源,底层的数据格式对上层应用程序而言完全透明,实现了系统功能的

分离,增强了可扩展性,并降低了风险。

2.3 Office 2003 信息检索服务概述

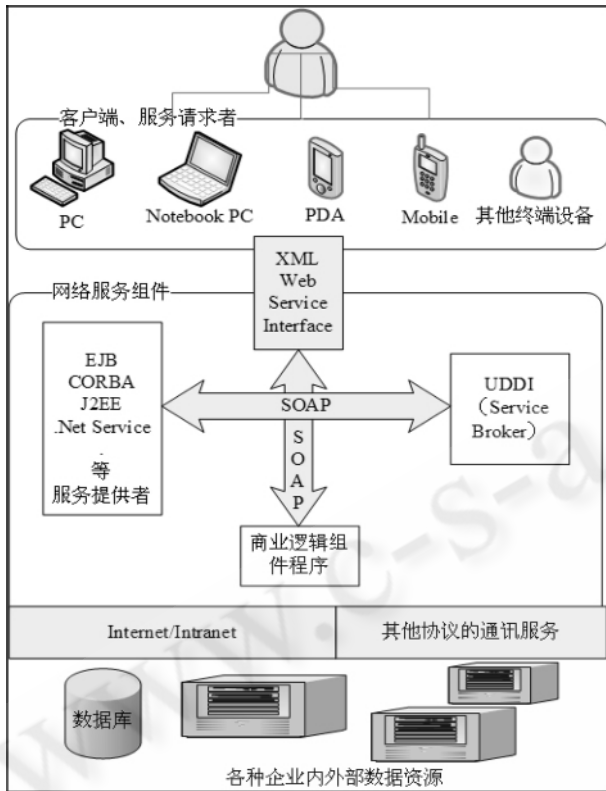


图 1 通用企业信息整合体系结构

Office 2003 中包含了一项非常有吸引力的功能—信息检索。从根本上说信息检索服务的后端就是一种网络服务,信息检索服务的前端是 Office 2003 中的所有组件程序 (Word、Excel、Outlook 等) 以及 IE 等浏览器,这是一种瘦客户端 (IE) 和智能客户端相结合的连接方式。信息检索的界面演示如图 2 所示。

该图显示了经过查询一个特定商品信息之后,在 Word 2003 中显示出来的页面,如果需要还可以把这些信息组织成 Excel 图表的格式,供用户快速的浏览和获取有用信息。

Office 2003 的信息检索服务能够提供丰富的界面定制功能,这些界面元素的背后都是标准的 XML 格式。这种灵活和统一的定义格式更加有利于提高用户的应用价值,实现信息的可视化,增强用户的体验和信

2.4 Office 2003 信息检索服务在企业信息化中的应用

Office 2003 信息检索服务具有连接网络服务接口的能力,因此 Office 应用程序可以方便地连接到网络服务接口上,利用网络服务整合不同系统并提供统一接口,实现对信息管理和控制的可视化,以图表和表格的方式显示数据,对用户来说是最熟悉和直观的方式,能够提高用户工作效率。并且该技术具有极高的处理效率,能够充分利用 XML 网络服务的平台无关性,实现分布式的网络信息共享。



图 2 Word 2003 信息检索服务界面演示

3 企业信息系统设计实例

3.1 背景介绍

在微软大中华区具有 2000 - 3000 名员工,原有的员工信息包含在以下的几种系统中:企业活动目录 (Active Directory, 该目录中包含有员工的一些基本信息)、Excel 表格、个人联系人的资料、原有的一个基于 Web 的微软全球员工信息查询系统。

目前员工的信息获取和管理主要有以下问题:

- (1) 信息分散,每个系统包含信息的一部分;
- (2) 信息保存的数据格式不统一,系统之间不能通信;
- (3) 用户使用和管理不方便,没有统一的界面和接口,无法实现正确信息的快速获取;
- (4) 信息更新和维护难度大,数据不同步问题严重。

基于以上所提到的问题,为了促进企业信息利用的自动化,设计和开发了一套供微软大中华区雇员使用的信息检索系统(EIS Enterprise Employees Information System)。上面提到的四个问题都是我们在本文开始介绍企业信息和数据管理中存在的问题,因此系统的设计和实现也与图 2 所介绍的解决方案总体保持一致。

3.2 系统设计

根据前面所介绍的利用 XML 网络服务和 Office 的信息检索技术的设计思路,员工信息检索系统(EIS)的体系结构如图 2 所示。它从用户视图和组件视图两个角度诠释了 EIS 的外部用户对象、内部系统构成及组成构件间的关联关系。

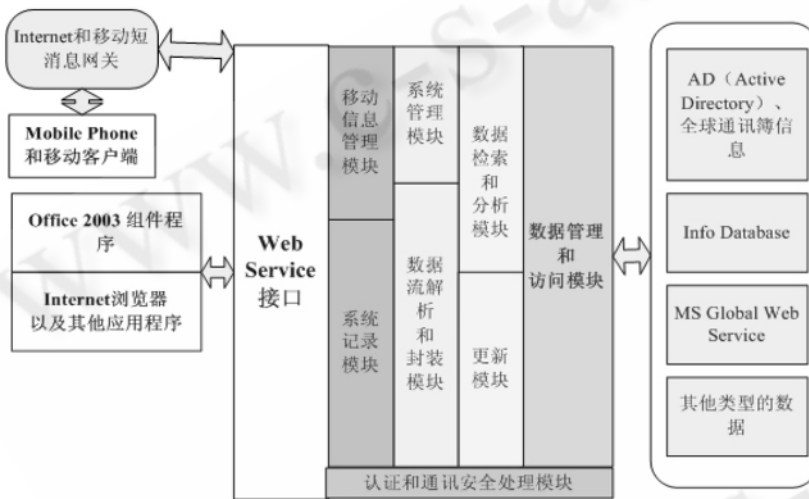


图 3 EIS 系统结构图

从用户视图来看,EIS 主要用户有两种:一般用户和管理员。在系统中也相应设立了两种角色:General User 和 Admin。这两种角色分别具有不同的权限,最重要的一点是普通用户不具有系统管理模块访问和使用的权限,这样的模块化设计和权限分配最大限度的保证了系统模块的独立性和系统实现的简便性。

从系统构成角度来看,EIS 系统是一种典型的三层结构。最后端数据层的数据提供者有:AD(活动目录)、MS 全球员工信息网络服务和 Database 等。

表示层主要由微软的智能客户端组成:Office 组件程序和 Smart Phone 移动客户端,同时提供了传统的 Web 浏览器访问方式。Office 智能客户端可以利用其

内置的信息检索技术,自动检索和分析特定封装的数据信息,并把信息和数据结果以 Word 文档、Excel 表格和图表、Access 数据库等方式显示出来,并且能够利用业务逻辑层的移动消息管理模块来发送移动短消息给手机用户。

业务逻辑层主要负责信息的查询、分析、管理、更新等功能,同时提供认证和信息安全处理、数据封装、数据访问等组件。业务逻辑层采用模块化设计,是为了有利于系统功能的分离和分层。

其中数据访问模块用于连接数据库和其他后台数据资源,它为所有上层模块和服务提供统一的数据访问接口。

数据检索和分析模块用于从数据库中查询特定的检索词,并分析和返回结果给数据封装模块。数据检索的算法封装在该模块中,几种不同的检索算法是独立的,并可以通过后台的配置文件进行配置。目前仅仅提供几种简单的检索算法,但是可以很容易扩充。

数据流解析和封装模块主要有两种功能:

- (1) 把查询结果封装成特定的 XML 数据格式发送给客户端;
- (2) 把客户端传送来的查询信息,解码成内部信息格式,供数据分析和检索模块或者更新模块使用。

更新模块可以完成用户或者管理员提交的信息更新功能。用户可以修改个人信息,管理员可以添加、删除和更新所有人信息。

安全管理模块用来完成用户身份认证以及数据加密传输的功能。在需要穿越防火墙的情况下,以及信息要求加密传输时该模块将起到至关重要的作用。

移动消息管理模块用于完成移动短消息的发送、接收和信息解析以及编码等功能。除此之外,还有系统管理和系统记录模块,完成一些系统管理配置和记录的功能。

以上所介绍的是系统各个模块的功能,从图 3 还可以看到,模块之间是有一定层次关系的,数据流的方向

(下转第 34 页)

(上接第 29 页)

向是从左到右的方式进行的,数据管理和访问模块是最底层模块,而移动信息管理和系统记录模块式最接近表示层的模块,数据流的传递基本按照图 3 所示的层次关系进行。只有认证和安全处理模块是贯穿整个业务逻辑层的。这样的层次划分是为了模块的独立性和扩展能力所设计的。

4 总结

随着网络技术向纵深发展,企业信息网络化是必然趋势,基于 Web 的分布式客户/服务器 Intranet 网络是当今企业信息化的最佳选择,而企业信息化的全面实现往往是多种信息化技术的综合:安全技术、可视化技术、搜索技术等。利用 Web Service 框架结构,并配合其他相关技术,能帮助企业抛开各类应用系统的对象体系、运行环境、开发语言等技术方面的束缚,打破地域的界限,建立稳定安全的电子信息传递通道,真正

实现高效的企业信息化环境。

参考文献

- 1 张义顺, Intranet - 企业信息化建设的成功之路当现代化工[J], 2001, 第 30 卷, 第一期。
- 2 杨建武、陈晓鹏, XML 相关标准综述, 计算机科学[J], 2002. 29(2)。
- 3 李鹏、张景, XML Web Service 在电子办公中的应用[J], 计算机工程与应用, 2004. 03。
- 4 柴晓路, Web 服务架构与开放互操作技术[M], 北京 清华大学出版社, 2002. 6. 56 - 58。
- 5 吕武, 企业信息化中的数据资源建设方法, 现代管理, 2004. 4。
- 6 Roger? Wolter XML Web Service 基础[EB/OL], <http://www.microsoft.com/china/msdn/archives/library/Dnwebsrv/html/webservbasics.asp>.