



陈寅生 编著

MCMS 2002

Microsoft
Content
Management

内容管理技术

Server 2002

國防工業出版社

<http://www.ndip.cn>

MCMS 2002 内容管理技术

陈寅生 编著

国防工业出版社

· 北京 ·

内 容 简 介

本书是国内为数不多的一本中文版的介绍 MCMS 2002 的使用和开发的书籍,也是一本讨论内容管理系统相关知识和企业内容管理技术应用的参考书。作者在较短的篇幅里总结了 20 世纪 90 年代以来以内容为主的 Web 站点在企业信息化建设中的经验和教训,并通过案例分析和开发实例来探寻一条使用 MCMS 2002 来高效快速地建立企业内容管理系统的道路。

本书从讲述内容管理系统的基本原理和基本方法开始,围绕微软公司推出的 Microsoft Content Management Server(MCMS)2002,具体分析了基于 MCMS 的内容站点的开发、内容创作和管理的实现技术与方法。本书的特点在于既给出了开发内容管理系统的设计思想,又介绍了企业内容站点的运营策略,并对部署内容管理系统的过程中可能遇到的实际问题加以分析。

本书是企业内容管理系统的开发人员和管理人员不可或缺的书籍,也可以作为其他以内容为主的 Web 站点设计者的参考书,本书也适合作为高等院校 Web 应用程序开发课程的参考书。

图书在版编目(CIP)数据

MCMS2002 内容管理技术/陈寅生编著. — 北京:国防工业出版社,2004.3
ISBN 7-118-03391-X

I. M... II. 陈... III. 超大规模集成电路
IV. TN47

中国版本图书馆 CIP 数据核字(2004)第 003196 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京奥隆印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 24 554 千字

2004 年 3 月第 1 版 2004 年 3 月北京第 1 次印刷

印数:1—4000 册 定价:33.00 元

(本书如有印装错误,我社负责调换)

前 言

本书是一本讨论内容管理系统相关知识和企业内容管理技术应用的参考书。我们在较短的篇幅里总结了 20 世纪 90 年代以来以内容为主的 Web 站点在企业信息化建设中的经验和教训,并通过案例分析和开发实例来探寻一条使用 MCMS 2002 来高效快速地建立企业内容管理系统的道路。在内容管理系统中,作为数字和信息的特殊结合体的内容被看作对象,它的一个重要属性就是它是可以被管理的。原始的信息只有按照一种可以被使用的形式封装之后才有可能成为内容,内容的价值还来自于内容的使用者对内容所进行的一系列的创建、保存、修改和共享的操作,目前市场上有多种内容管理的解决方法,本书以微软公司推出的 Microsoft Content Management Server 2002 (MCMS 2002)为例介绍开发、部署和运行内容管理服务的原则和方法。

MCMS 2002 是微软公司推出的一个内容管理产品,它允许企业快速高效地建立、部署并维护高度动态化 Internet、企业内部网络及企业外部网络 Web 站点的企业级 Web 内容管理系统。它是一个用于管理企业级 Web 内容的全面系统,允许内容提供者创建、管理并发布其自己的内容,同时允许 IT 部门快速部署具备伸缩能力的动态站点。MCMS 2002 具备动态内容提供的功能,企业级 Web 站点可以针对多种访问者、设备及目的提供动态内容。相关内容的个性化特性能够确保为潜在客户、客户、企业员工及商务合作伙伴提供积极的用户体验。MCMS 2002 具备快速开发与部署内容管理系统的功能,可以用来部署企业级可伸缩性动态内容 Web 站点,并实现比其他任何 Web 内容管理产品都要明显加快的投资回报。

MCMS 2002 允许站点管理人员使用一个创建目录的标准化方法为 Web 站点的开发定义 workflow 结构。规划设计小组构建了网站的结构,资源开发人员创造图形和其他支持资源,内容是作者的工作,编辑和校对者改进内容。MCMS 负责安全和网站创建的其他方面以及是如何提供内容服务的。

MCMS 2002 的框架

了解 MCMS 2002 的关键的组件以及其扮演的角色是计划一个部署内容管理服务的步骤。MCMS 包括一些核心组件,这些组件共同用于为网站开发创建 workflow 结构以及网站的内容访问通路。MCMS 依赖于 Internet 信息服务器作为 Web 服务器,不仅仅用于客户对内容的请求还包括编辑和网站的许多方面。服务器端 Web Author 组件提供了在开发者和服务器之间连接。Web Author 使开发者和内容创建者可以使用 IE 来新建和编辑内容而无需一个专门应用程序。

MCMS Site Manager 是一个 Windows Form 的客户端应用软件,网站管理员使用它来创建站点的安全框架,这里包括定义开发任务、创建容器和设置访问控制权限。设计者

和开发者使用 Site Manager 来创建模板, 在这个模板中作者用内容进行填充。资源开发者使用 Site Manager 来管理被模板设计者和作者使用的资源。

MCMS 应用软件安装在一个或多个服务器上, 在网站数据库和客户以及开发组之间扮演中间人的角色。MCMS 服务器根据来自于客户端的请求自动建立来自于被储存在内容库中的数据, MCMS 也通过缓冲内容来改进性能。

在站点结构的后端需要解决如何保存内容, 即一个 SQL 数据库和保存了站点所有内容的对象模板, 包括模板、资源图表、网页和相关内容。MCMS 将数据库技术和 Windows 文件系统结合在一起, 在 Resource 资源库中保存任何类型的数据, 包括文档、可执行文件、多媒体等。

此外有两个组件有助于开发者和管理员创建和配置站点。网站部署文件 SDO 允许从一个 MCMS 服务器到另一个以一个 XML 包的形式输入一个站点; 该应用程序可以保证源服务器到目标服务器的内容同步, 在整个传输过程中将在源数据上加锁, 这样确保了服务器之间的数据完整和一致。

另一个就是 Site Stager, 管理员使用它为没有 MCMS 的 Web 服务器以 HTML 格式输出站点。Site Stager 创建了站点以及所有其内容的一个 HTML 快照。任何 Web 服务器可以处理结果内容, 包括非 IIS 服务器。Site Stager 允许网站管理员创建分段传输说明, 定义输出内容以及输出时间。我们可以设定自动输出, 或这也可以手工完成整个过程, 这依赖于站点的需要。我们可以使用 Site Stager 在多个位置对该站点进行备份。

本书的结构

从上面的介绍中可以发现使用 MCMS 开发与部署内容管理服务牵涉到了整个微软体系结构的方方面面, 本书的一大特点就是对这些方面的内容加以充分的关注。我们还将讨论内容管理系统相关知识和企业内容管理技术, 通过案例分析和开发实例来探寻一条使用 MCMS 2002 来高效快速地建立企业内容管理系统的道路。本书一共分为 10 章。

第 1 章是对内容和内容管理的概述。内容是什么类型的数字信息的结合体, 是一个比数据、文档和信息更广泛的概念, 是对各种结构化数据、非结构化文档。信息的聚合。内容可以表现为文本、图形图像、Web 页面、业务文档、数据库表单、视频、声音文件等存在形式。而内容管理就是实现各种数字信息内容的创建、储存、分享、应用、更新操作, 并在企业、个人、组织、业务、战略等诸方面产生价值, 从而将内容中蕴含的数字财富挖掘出来的系统工程。

第 2 章是对内容管理系统的介绍。当我们明显需要实现内容管理技术的时候, 购买它实际上是规划与部署企业内容管理系统过程中最容易的一部分。而最难的部分是确定我们需要让 CMS 解决什么问题, 还有我们是否确实需要一个内容管理系统。在该章中我们将以一个从内容型网站发展而来的企业内容站点作为例子, 来分析需求是如何产生的, 并告诉你为什么每个成功的 CMS 工程都起始于可靠的商业策略。

第 3 章的内容可以被作为是对 MCMS 2002 的一个快速浏览。MCMS 2002 是一个用于管理企业级 Web 内容的全面系统, 允许内容提供者创建、管理并发布其自己的内容, 同时允许 IT 部门快速部署具备伸缩能力的动态站点。

第 4 章分析内容管理系统对于企业信息化建设的贡献, 我们需要把内容管理集成到

企业信息化系统之中。

在第 5 章中向读者介绍了如何配置一个 MCMS 2002 开发环境用来评估与开发基于 MCMS 2002 的应用,包括安装 MCMS 2002 所需的一系列软、硬件需求,以及如何在企业中一台或多台服务器上安装与配置内容管理服务。

第 6 章我们将关注基于 MCMS 的内容管理服务的规划与部署,尤其是对于大型的系统而言,充分的规划工作必不可少。将首先对部署工作的目标体系结构加以分析,然后提出站点部署的蓝图,确定项目的范围,确定开发时间,并建立一系列的评估标准。MCMS 可以被部署到多种体系结构之中,我们还将着重讨论集群为运行可靠的内容管理服务提供的保障,包括故障转移和负载均衡特性。在该章中还将把内容管理系统与基于 Microsoft Commerce Server 的电子商务系统集成在一起,这个实例可以被看作是第 4 章中所涉及内容的继续。

在第 7 章中将介绍了 MCMS 2001 站点迁移工作需要考虑的一系列问题,并结合一个传统的 ASP 程序的例子来与读者一同完成迁移工作。

第 8 章是对 MCMS 的深入介绍,将了解基于 MCMS 的 Web 站点是如何使用 MCMS 内容管理服务的,包括站点内容的创建和发布和站点内容的管理,另外还有一个方面对于内容管理系统非常重要,这就是 XML 的使用。我们将讨论什么是 XML 技术,为什么要使用 XML 技术。

优化 MCMS 站点的性能有多种方法,在第 9 章中讨论开发人员如何可以通过优化 ASP.NET 应用程序的性能来保证 MCMS 站点的性能,这是最经济和简便的方法。在本书中我们一直非常关注站点的安全性,对于一个企业来说一个不安全的,容易受到外部攻击的内容管理系统毫无价值,MCMS 充分集成了 Windows 的安全性服务,可以将这种安全特征划分为多个层次,包括平台安全性,代码安全性和服务安全性。在该章中将分析 .NET 架构提供的安全服务,以及开发可靠的 ASP.NET 应用程序。

第 10 章的内容将体现 MCMS 2002 的一个重要的特点,就是与 VS.NET 的紧密集成,这个特点对于站点开发人员特别有意义。在 MCMS 2002 之前站点开发人员所使用的的是一个非专业的开发环境,有的人会使用 FrontPage,而另外一些开发人员直接使用 Notepad 作为工具。现在站点开发人员能够在 VS.NET 中直接建立基于 MCMS 的 Web 站点项目,VS.NET 为开发人员提供了一个集成的开发环境,在该章中我们将了解如何使用 VS.NET 开发 MCMS 应用。

MCMS 为开发人员提供了一个基于 .NET 框架的对象模型 (MCMS Object Model),它分为 MCMS.NET 类库和 COM 对象模型两大类,本书的附录可以供使用 MCMS 对象模型来开发内容站点的开发人员参考。

网上资源

在 Internet 上可以找到一些非常有用的内容来帮助学习和掌握 MCMS 2002。微软公司为 MCMS 2002 建立了一个内容丰富的 Web 站点,并且这个站点本身就是一个基于 MCMS 的内容站点的示例,该站点的位置在 <http://www.microsoft.com/cms>。MCMS 的用户还可以在该站点上下载最新的 MCMS 2002 的升级程序,对于使用 VS.NET 2003 作为开发工具的 MCMS 2002 开发人员,可以在站点上下载 MCMS 2002 Service Pack 1。该升

级程序可以使开发人员能够在 .NET Framework 1.1 环境之下使用 MCMS 对象模型。

另外一个需要关注的地方是 MSDN，这是一个为开发人员提供服务的站点，在那里可以看到对于 MCMS 功能的大量深入的探讨，并且有示例代码可以下载，MSDN 的地址为 <http://www.microsoft.com/msdn>。对于 MCMS 用户来讲微软公司的 TechNET 也是一个不错的去处，TechNET 的目标用户是企业中的知识工作者，里面提供了很多有益的使用信息和技巧提示，其地址为 <http://www.microsoft.com/Technet>，这些站点的内容的更新非常及时。

编者

2004.1

目 录

第 1 章 管理有价值的内容.....	1
1.1 挖掘内容中蕴含的价值.....	1
1.1.1 定义需要管理的内容.....	1
1.1.2 复用为内容增值.....	4
1.1.3 内容的质量.....	8
1.2 什么是内容管理.....	12
1.2.1 内容管理的迫切性.....	12
1.2.2 内容管理的内涵.....	15
1.2.3 内容管理的发展趋势.....	18
1.2.4 内容管理与内容服务.....	20
第 2 章 内容管理系统概述.....	22
2.1 从内容站点到企业内容站点.....	22
2.1.1 开发一个内容站点.....	22
2.1.2 管理站点资源.....	30
2.2 内容管理系统的需求分析.....	35
2.2.1 开发与部署内容管理系统的时机.....	35
2.2.2 不同行业对于内容管理系统的需求.....	36
2.3 开发内容管理系统.....	38
2.3.1 内容管理系统的基本特征.....	38
2.3.2 内容管理系统的实施.....	40
2.3.3 内容管理系统的架构.....	41
2.4 日趋成熟的内容管理系统.....	44
2.4.1 内容管理实施工具概述.....	44
2.4.2 选择内容管理系统的原则.....	45
2.4.3 主流内容管理软件.....	46
第 3 章 MCMS 2002 入门.....	49
3.1 微软企业内容管理技术概览.....	49
3.1.1 系统架构.....	49
3.1.2 MCMS 2002 的内容管理特性.....	52
3.1.3 使用 MCMS 2002 进行内容管理的优点.....	54
3.1.4 MCMS 2002 中的用户类别.....	55
3.2 MCMS 2002 的功能概览.....	56

3.2.1	内容提供与发布	56
3.2.2	MCMS 内容站点的管理	59
3.2.3	站点开发	61
3.3	使用 MCMS 2002 的效益	63
3.3.1	MCMS 2002 降低了内容站点的业务成本	63
3.4.2	使用 MCMS 2002 的商业收益	65
第 4 章	在企业信息化系统中集成内容管理	68
4.1	内容管理与企业信息化	68
4.1.1	企业信息化系统的实施关键	68
4.1.2	内容管理与知识管理	70
4.1.3	内容管理与客户关系管理	73
4.1.4	内容管理与企业文化的转变	76
4.2	微软企业服务器产品概述	79
4.2.1	完整的企业解决方案	80
4.2.2	用于部署和操作的 IT 基础结构	81
4.2.3	用于设计和开发的应用程序基础结构	84
4.2.4	用于通信和协作的信息基础结构	91
第 5 章	配置 MCMS 2002 开发环境	100
5.1	安装 MCMS 2002 的准备工作	100
5.1.1	组件升级	102
5.1.2	增强 Internet 信息服务的安全性	113
5.1.3	使用 Visual Studio.NET 2002 作为开发工具	123
5.1.4	兼容性	126
5.2	安装 MCMS 2002 内容管理服务	128
5.2.1	配置 Windows 安装服务	128
5.2.2	为 MCMS 创建系统账户	129
5.2.3	在 SQL Server 中创建 MCMS 数据库	132
5.2.4	配置 Internet 信息服务	137
5.2.5	安装 MCMS 2002 的组件	140
5.2.6	使用 DCA 来配置数据库	144
5.3	MCMS 2002 中的内容管理与创作组件	149
5.3.1	安装 Authoring Connector	149
5.3.2	激活 Web Author	151
第 6 章	规划与部署内容管理服务	153
6.1	规划一个基于 MCMS 2002 的内容站点	153
6.1.1	概述	153
6.1.2	在微软体系结构中进行站点部署	154
6.1.3	规划内容站点蓝图	160
6.1.4	为开发基于 MCMS 2002 的内容站点制定规划	164

6.1.5 制定站点的配置计划	168
6.2 部署 MCMS 2002	168
6.2.1 概述	168
6.2.2 使用集群来确保服务可靠性	170
6.2.3 将 MCMS 2002 部署到集群之上	176
6.3 面向电子商务站点的内容管理	182
6.3.1 集成架构	182
6.3.2 集成工作的物理部署	184
6.3.3 部署方案	186
6.4 建立使用 MCMS 2002 的内容站点示例	190
第 7 章 迁移到 MCMS 2002 与 ASP.NET	193
7.1 升级到 ASP.NET 需要考虑的问题	193
7.1.1 核心 API 的变化	193
7.1.2 结构变化	194
7.1.3 应用程序配置的变化	196
7.1.4 状态管理	196
7.1.5 ASP.NET 中的安全性	197
7.1.6 数据访问	198
7.2 升级到 ASP.NET	200
7.2.1 在同一个 Web 服务器中运行 ASP 和 ASP.NET	200
7.2.2 VB.NET 语法上的改变	200
7.2.3 回顾 ASP 应用程序	202
7.2.4 将 ASP 应用程序升级到 ASP.NET	205
7.3 将基于 MCMS 2001 的站点迁移到 MCMS 2002	210
7.3.1 概述	210
7.3.2 规划迁移工作	213
7.3.3 迁移到使用 ASP.NET 的站点	215
7.3.5 让站点中的 ASP.NET 与 ASP 混合使用	216
第 8 章 MCMS 2002 中的内容管理技术	218
8.1 站点内容的创作与发布	218
8.1.1 使用 Web Author 创作内容	218
8.1.2 从 Microsoft Word 中发布内容	232
8.2 管理基于 MCMS 的内容站点	238
8.2.1 服务器配置程序 SCA	239
8.2.2 创建和维护站点的栏目	241
8.2.3 规划内容管理工作流	246
8.3 使用 XML 技术表示内容	250
8.3.1 用 XML 来表示数据	250
8.3.2 使用 XML 来表示内容的优点	254

8.3.3	在内容站点中使用 XML 技术	255
8.3.4	在 MCMS 2002 中使用 XML 技术	259
8.3.5	利用 XML 技术为内容编制索引	259
8.4	使用 MCMS 2002 部署国际化站点	267
8.4.1	内容站点的国际化与本地化	267
8.4.2	国际化站点的规划与实施	277
8.4.3	MCMS 2002 中的多语言支持	279
第 9 章	MCMS 2002 的性能和安全性	281
9.1	优化 ASP.NET 程序	281
9.1.1	管理 ASP.NET 应用程序	281
9.1.2	监视 ASP.NET 应用程序	282
9.1.3	工作负荷分配和 ASP.NET 应用程序	283
9.2	MCMS 2002 的站点维护	285
9.3	MCMS 2002 中的用户身份验证机制	287
9.3.1	MCMS 中与安全性相关的系统结构	287
9.3.2	MCMS 的用户身份验证机制	288
9.3.3	需要关注的安全设置	291
9.4	可信赖的 MCMS 2002: 站点安全性指南	292
9.4.1	Microsoft .NET 的安全性概览	292
9.4.2	构建安全的 ASP.NET 应用程序	295
第 10 章	基于 MCMS 2002 的应用开发	301
10.1	使用 VS.NET 作为开发工具	301
10.1.1	使用 VS.NET 集成开发环境	302
10.1.2	.NET 应用程序的体系结构	310
10.2	在 VS.NET 中建立 MCMS 工程	315
10.3	实现内容模板开发中的源代码控制	319
10.4	MCMS 2002 中的集成技术	326
10.4.1	第三方验证系统	326
10.4.2	开发自定义内容占位符控件	331
附录 A	MCMS.NET 类库参考	338
附录 B	MCMS 2002 COM 对象参考	351

第 1 章 管理有价值的内容

在知识经济时代,取得商业竞争优势的关键在于对市场变化的准确把握,以及利用这些信息迅速制定科学决策的过程。内容管理的主要目的是把企业的信息资本演化为生产力。内容是什么类型的数字信息的结合体,是一个比数据、文档和信息更广泛的概念,是对各种结构化数据、非结构化文档、信息的聚合。内容可以表现为文本、图形图像、Web 页面、业务文档、数据库表单、视频、声音文件等存在形式。而内容管理就是实现各种数字信息内容的创建、储存、分享、应用、更新操作,并在企业个人、组织、业务、战略等诸方面产生价值,从而将内容中蕴含的数字财富挖掘出来的系统工程。

1.1 挖掘内容中蕴含的价值

对企业运营来说,真正有价值的内容指的是把那些散乱的数据和无序的信息整合后形成的有组织、有价值的知识存在。内容管理就是通过对一个企业、部门或者 Web 站点的所有信息资源的重整和共享、对现有信息的管理和协同工作、对隐含的数据的挖掘,并利用软件解决方案而实现的组织自身管理与对外交流的方法。内容管理又是一种行为,是利用组织内外的知识工作者的智力资产直接产生实际的结果;知识管理还是一个过程,它通过使用包含在文档、系统中或其他形式的信息,或者个人的专业技能,使得个人和团队能够通过一定的系统进行协作,有效地利用知识来产生竞争和商业上的优势。

1.1.1 定义需要管理的内容

随着信息化进程的推进,越来越多的政府部门、企事业单位和各种非政府组织都建立起自己的 Web 站点,进行电子政务或者电子商务的活动。在上网的过程中,无论是通过互联网发布公司复杂的产品和服务目录,还是提供客户支持信息,或者进行企业内部事务的整合,都不可避免涉及到内容的发布与管理工作。内容管理是与电子商务紧密相关的,而电子商务,不管是企业对企业的,还是企业对客户的,它们之间都需要交换数据,因此内容成为一个重要的推动力。信息包括了结构化信息以及非结构化数据,数据库中的数据是被格式化的,能够被计算机所管理的,而例如企业员工之间的电子邮件中的信息就是非格式化的。

首先,需要明确的一个重要概念就是什么样的内容需要加以管理,这是内容管理的前提。内容是什么类型的数字信息的结合体,是一个比数据、文档和信息更广泛的概念,是对各种结构化数据、非结构化文档、信息的聚合。内容可以表现为文本、图形图像、Web 页面、业务文档、数据库表单、视频、声音文件等存在形式。

从起源的角度看,内容一词首先起源于出版业。书报、杂志和唱片影带里的元素就

叫做内容，所以早期对于内容的定义偏向出版产品中的元数据。这些内容的储存多以流形式保存在数据库之中，或者以一般数据文件方式储存。早期的内容对象从制作编辑到成品储存都要经过的或多或少的加工过程一般都是简单的流水线方式的流程，分叉多线的并行操作并不多见。不同类型的内容元素的处理流程相差甚远，很难结合在一起。对于元数据的描述，也主要是以数据属性的方式来记录，比如内容的作者、创作日期、关键词、媒体种类、版权等。

随着网络和信息技术的普遍使用，企业内部信息的流通速度大大加快，文件与资料的数量也快速增加，除了传统的媒体形式以外，数字化的文档比如电子邮件、Word 文档、网上讨论、实时简讯等内容在企业的日常事务中发挥着越来越重要的作用，于是内容就由原来的出版媒体中的狭义的内容扩大成企业内部全部资料信息的广义定义。在今天，不管何种形式的内容都可以以 XML 标记的方式记录，并通过 XML 来达到跨组织发布和资源共享的目标。

企业内容管理是一个全新的研究与应用范畴，目前对于“什么是内容”还没有非常明确的定义。本书中所关注的内容，首先是多种类型的数字信息的结合体，内容需要具备可数字化的特征，也只有数字化以后的内容才能够被应用到各种内容管理系统之中。内容的存在形式可以是文本、图形图像、Web 页面、业务文档、数据库表单、视频、声音文件等，从这个角度说，内容是一个比数据、文档和信息更抽象的概念。再次，内容是对各种结构化数据、非结构化文档、信息的聚合，这在某种程度上也包含了知识。按照这个定义，可以对现有的企业信息加以分析，从而得出在现实环境中，需要关注的内容通常包括如下几个方面：

(1) 企业数据库中保存的大量业务数据。

(2) 通过后端办公处理系统发布并且能够被客户、合作伙伴和客户服务代表获得的业务文档，如合同和发票等的各种表格。

(3) 企业员工与商业伙伴之间的电子邮件。

(4) 多媒体内容，如音频和视频文件。

(5) 针对每一名用户所进行的个性化设计的 Web 内容。

在进一步了解内容管理技术之前，首先要区分数字 (Data)、信息 (Information) 和内容 (Content) 这三者之间的关系。数字是指对于使用者具有意义的信息单元，在这里，其范围仅仅局限于计算机所要处理的数据。在早期的计算机应用情景之下，比如最早的批处理系统的使用者就需要将他们编制的程序和数据记录在用打孔表示的卡片上，这时候的数字是相对孤立的，按照二进制的位 (bit) 来表示的，而一台计算机的内存所能处理的数据的量非常有限。随着存储介质的更新换代，人们可以以更加宽裕的方式来保存信息，人们开发了数据库来保存大量的数据，但是问题依然存在。因为在这个阶段，计算机处理的信息仍然是以数字的形式抽象出来的，换句话说，数字依然是信息的主要表现形式。随着计算机在商业领域的普及，数字多媒体技术的出现，以及大容量存储介质 (DVD) 的广泛使用，真实的信息不用被转换为抽象的数字符号来保存和使用，而能够以最自然的方式被记录和再现。比如人们需要获取关于某个亲友的信息，如果处于数据库的时代，那么他的查询结果将返回一系列的字符信息、年龄、性别等。而现在，得到的就可能不仅仅是上述的信息，还可能获得这个人的一幅照片、一段视频图像或与这个

人相关的一些 Web 链接。

不仅如此,人们已经对于从 Internet 上获得这样的内容习以为常了,现在用户还可以以其他任何的方式来获取内容,如通过 PDA、手机,甚至一台连接了无线网络的冰箱。随着企业信息化水平的提高,人们已经越来越关注如何处理内容。简单地说,内容也是一种信息,它与数据最大的不同在于,内容不仅表示了一定的意义,同时它还具备逻辑关系。这里所谓的逻辑关系指的是组成内容的元素之间的前后相关的结构。所以说,数字是内容的一种表现形式,但是内容却不仅仅是数字。

数据是人们收集、整理并保存在文件或者数据库中的记录,而信息的概念就更加宽泛了,它不仅包括了有用的数据、各种多媒体材料,也包括庞杂而无意义的材料。在这里将信息的范围限制在有用信息的范畴,信息可以被定义为人们所进行的一切交流行为的记录,它包括文字材料,比如小说、新闻甚至文字广告,以及声音、图像、视频和各种被保存在电脑之中的 PowerPoint 幻灯片、Excel 表格等。因此,与数据相比,人们花费在信息上的时间与成本要比数字多得多。比如编制一个 Excel 的电子表格,不仅要获得输入表格中的各种数据,还要定义表格的结构和样式,如果有必要的话,还要添加一些自定义的函数用来处理数据。当然这种信息也存在着问题,比如这张 Excel 的表格,如果人们只是获取到表格中的某些行的内容,那么对于他们来说,这些信息是毫无用处的,因为信息被破坏了,这样的情况在商业社会里经常发生。流信息的概念,不仅被用来说明诸如声音,视频这样的多媒体对象,也可以从一个侧面来反映出信息的尴尬境地。所以这样的信息同样不是内容。

在本书中将内容定义为对象,作为对象的一个重要属性就是它是可以被管理的。原始的信息只有按照一种可以被使用的形式封装之后才有可能成为内容,内容的价值还来自于内容的使用者对内容所进行的一系列的创建、保存、修改和共享的操作。总的来看,内容应该具备下列特征:

1. 内容是数字和信息的结合体

内容不是单纯的数字,内容不仅是数字的一个容器,同时也包括了对于数字的分析、过滤和钻取。信息也不足以构成内容,这是因为除了信息中的元素以外,人们对于信息的感知更多地来自于自己对于客观世界的认识。内容不仅可以将信息表达出来,同时通过各种技术(比如自然语言)理解和搜索技术获得与信息相关的其他信息,并能够以多媒体的方式展现出来。所以说,内容是数字和信息的结合体。

2. 内容有用性

内容的有用性体现了内容的价值,对于内容的处理,需要耗费大量的人力和物力,如果内容是无用的,那么这部分的投入就会白白浪费掉。内容还是可以增值的,纯粹的信息或者是数据被加工为内容,这种加工不仅仅是将不同的数据结合在一起,也可能包括对于内容相关信息的搜集,或者是对于组成内容的数据的分析比如数字挖掘技术。

3. 内容具备生命周期

内容是被用户所创建的,在为内容设置信息和数据的同时,用户还需要为内容标定状态。这是因为作为企业知识链上的重要载体,内容需要被计算机所处理,内容还需要被企业内部和外部的组织或者个人共享。因此内容的状态就显得非常重要了,一般来说,内容的状态可以划分为三个阶段,即新内容、已发布的内容和已删除的内容。新内容总

是应该被给予更多的关注，已经被删除的内容也仍然有其价值，同样需要合适的处理。

4. 内容具备格式

内容的格式有两层含义，首先是内容的保存格式，即内容在存储介质的保存和传播的格式、计算机只能处理 0 和 1，所以内容的保存格式仍然为二进制格式。但内容的表现格式是多种多样的，这也是内容的价值的重要体现，用一个内容可以包括多种的媒体形势，文字、图像和声音可以在内容中获得统一。

5. 内容具备结构

内容的结构是指组成内容的各个对象之间的相互关系，以及如何让人和计算机两者都能够理解内容的意义。描述内容的结构曾经是一件非常困难的事情，不过这种情况已经随着 XML 的使用得到了很好的改善，XML 是真正的企业标准，并且可以扩展，它能够描述各种类型的对象。

1.1.2 复用为内容增值

提高现有内容的使用价值是企业内容管理的重要目标，内容复用在这个方面起到关键的作用。系统化的复用策略保证了内容的前后一致性与连贯性，降低了内容创作的成本，还能够方便地进行内容的重新装配。内容复用不仅是在现有的内容的基础上重复使用，而是将按照对象的方法将内容素材链接起来。

复用并不是一个全新的策略，在信息技术领域的其他部门已经广泛地使用了这一策略，并带来了可观的效益。复用的策略最早来自传统的制造业，标准化的零件可以被作为组件在各种不同的产品中使用。早在 30 年之前，软件公司为了面对日益严重的“软件危机”开始研究基于组件的软件开发技术。在 20 世纪 90 年代，开发这种技术被广泛地接受，面向组件的开发模式逐步代替了面向对象的软件开发方法。具备一定功能的代码被封装在一起，通过定义好的接口与软件的其他部分进行通信。软件行业使用组件技术来保证产品的质量，降低开发成本，提高软件的发布日期。出版业是使用复用的另一个例子，由于多媒体技术的发展，出版商不仅需要以传统的方式来发行图书，同是也需要在 CD 上提供电子版本。Internet 也是内容复用的重要使用领域，站点的横幅广告在很早的时候就通过复用技术来进行管理。由于网站规模的扩大，网站的内容编辑人员为了保证站点所提供的信息的前后一致性付出了需要的劳动，一个好的内容复用策略能够显著地降低这方面的成本。

内容复用是一种把已有的内容作为素材，在此基础上通过原始材料的引用和修改来创作新的内容的一种方法。可以用来重复使用的内容包括文字素材、图片、表格和其他的内容形式等。其中文字素材的复用是最简单也最普通的一种复用，可以选择一段文字材料，然后通过复制和粘贴的方式将它们引用到自己的文章当中，直接使用或者是加以修改。以前，图片、声音等多媒体素材都是作为一个实体进行复用的，在 XML 技术的帮助下，也能够按照使用文字材料的方式来使用多媒体信息，比如作为 XML 扩展的 SVG 格式能够使得图片可以携带更多的信息，除了图片本身以外，SVG 还定义了图片的编号和简短说明等属性。

这种复用的方式在开始阶段能够发挥很好的作用，但是如果需要对引用的内容进行多次修改的时候就会发生问题。并不是每个作者都能够记住文章中的哪部分是引用的，

是从哪里引用的。将这些内容都找出来并且加以编辑是一件非常浪费时间的事情，而且很多时候一些需要修改的引文会被遗漏掉。从一个企业的角度来看，这样的情况时常发生并且可能带来严重的后果，同样的内容中的数据或者其他的信息将会出现前后不一致的情况，甚至到最后正确的数据或是信息会被这种错误所淹没。

所以，内容复用不仅是在现有的内容的基础上的简单重复使用，而是将内容链接起来加以管理，这是内容管理系统的一个基本原则。被引用的内容将以对象的形式被插入到正在编辑的文章当中，但是对于该内容的任何修改都是对原始材料进行的，并能够立即体现到所有使用了这个内容对象的其他文章当中。比如需要在 Word 文件中插入一幅图片，但是这幅图片中的内容很可能被修改，为了避免多次修改，可以以链接的方式将这幅图片插入到 Word 文档之中，如图 1-1 所示。对于图片所作的任何修改都能够及时地更新到这个 Word 文档中。这样的例子很多，比如用户也可以使用外部引用的方式直接引用 Internet 上的数据或者一个 Excel 表格中的数据，而无需关注这些数据是否会在不久的将来发生变化。

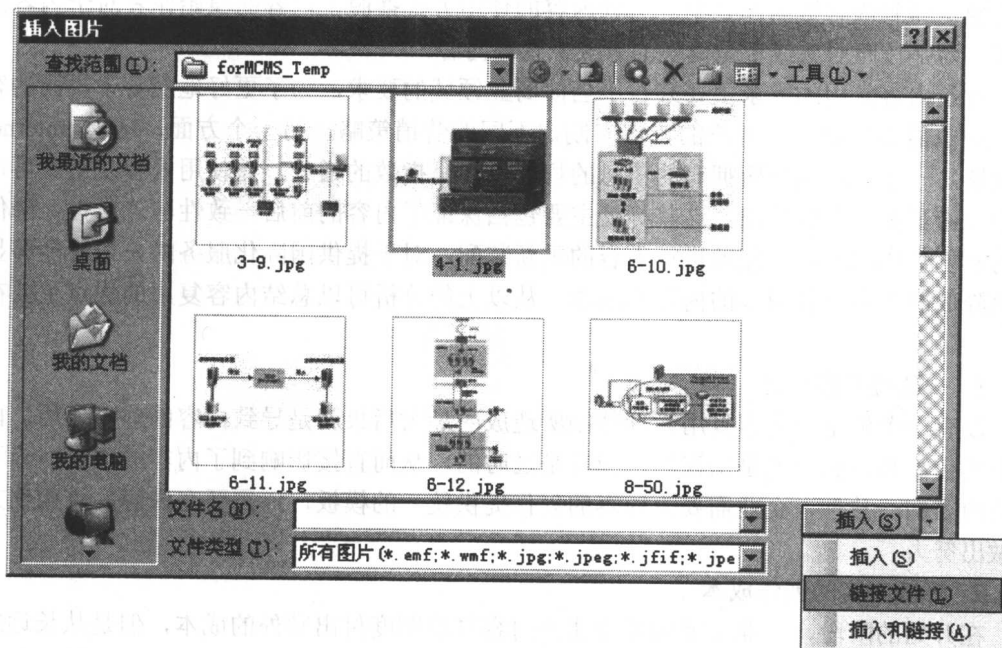


图 1-1 在 Word 中插入一个对于图片的引用

内容复用也被称为单一内容源头 (Single Sourcing) 技术，被保存在内容库中的内容素材可以被重复地在多种媒体和多个文档中使用。在知识管理领域，内容复用经历了简单复用、静态复用和动态装配三个发展阶段。

简单复用是指对于同一个内容素材根据承载内容的媒体的特点来进行复用。在这种复用形式中，对于内容的数据或者信息本身并不进行编辑和修改，而只是从内容源中选择适合特点媒体提点的材料。比如企业将某个产品的介绍性材料放在 Internet 上，并配以图片或者其他多媒体内容，而在这个产品的使用手册上需要的是文字性的材料，这些内容能够对产品的使用起到指导作用而不仅仅是加以介绍。在这种方式下，对于内容本身

及其表现形式基本上不做或者只做很少的改动,这是因为没有一种机制可以保证被修改的内容被及时更新到使用了用以前内容的其他材料之中,对于引用的信息或者对数据所作的修改,不论对内容源本身还是引用的材料,都会造成内容的不一致性。所以这种方式一般适合于不会进行频繁修改的内容。

内容复用的第二个阶段是静态复用,这种方式下内容源中的素材可以根据使用的需要进行必要的修改,原始作者在编写内容的时候就预先定义了可以进行修改的元素,或者为同一份内容根据使用特点编制不同的版本。其他人在引用的时候就可以对可以修改的部分加以编辑,或者是选择一个需要的版本。在这种方式下保证了内容的关键部分的一致性,并且在使用的时候能够对其余部分加以定义,但是它的灵活性仍然是有限的。

最后的一种方式动态装配。在动态装配模式下,内容并不是静态地保存在文档之中或者是以文件的方式存储在硬盘之上,而是保存在数据库之中。内容在需要使用的時候才被按照要求并根据一定的形式结合在一起。在这种方式下,内容是以对象的形式被保存和编辑的。用户对于内容中的信息和数据的修改也可以根据可以配置的用户权限被动态地更新到内容数据库之中。这种内容的使用方式摆脱了内容的原始作者对于内容的束缚,而是按照对象或者是模块的方式来管理内容。

对于内容复用的需求主要来自于当前商业领域的要求,为了更好地服务于不同的客户,企业需要针对各个客户的特点来制定不同的营销策略,另一个方面,随着 Internet 的发展,一个企业本身所拥有的信息的规模也在呈指数的增长,而使用内容复用就可以为企业带了多方面的效益。这些好处主要包括保证了内容的前后一致性与连贯性,降低了内容创作的成本,方便地进行内容的重新装配,对于提供国际化服务的公司还能够显著地降低制作多语言版本的内容的成本。从以上的分析可以总结内容复用的优点主要在于:

1. 一致性与连贯性

没有一个健全的内容复用管理体制所造成的最大后果就是导致内容的前后矛盾,内容中的信息和数据可能是错误的,或者是过时的,从而直接影响到了内容的质量。为了保证内容的一致性,企业需要为内容的创作提供统一的模板,还需要在内容的结构化方面做出努力。

2. 降低内容的创作成本

在开始的阶段,企业可能需要为建立内容复用制度付出额外的成本,但是从长远的角度来看,这笔花费是绝对值得的。通过搜索企业的知识库或者是门户网站,员工和客户都可以从中挖掘和获取有价值的内容,使用它们能显著地提高工作效率。对于内容创作人员来讲,他们可以在已有的内容中获取帮助,而不是每次都需要从零开始,并且在使用内容的时候不需要花费大量的时间来核对数据的正确性,也不会对同一主题数据的不同版本而感到迷惑。当引用的内容发生改变的时候,内容复用系统会自动地进行更新,而不要求人工的干预。内容复用还可以保证企业的客户通过 Internet 获取的信息都是最新的,这样在内容发生改变的时候企业就不再需要为了通知客户而花费额外的费用。

3. 内容可以进行重新装配

这是因为可以复用的内容是按照模块的方式动态生成的。信息的内容在发生改变的时候,这种方式可以很方便地实现内容的重新组装和发布。创作人员可以改变内容模块