

# 数据仓库和 OLAP 技术在政府网站评估中的应用

寿志勤, 刘 波

(合肥工业大学 管理学院, 安徽 合肥 230009)

**摘 要:** 政府网站评估是电子政务建设的重要组成部分, 持续的评估工作积累了大量的数据。如何利用这些宝贵的信息资源为网站评估管理与决策服务, 是政府网站评估深入发展的重要研究课题。将数据仓库和 OLAP 技术引入到政府网站评估分析系统, 建立了网站评估多维数据立方体, 并通过 OLAP 多维结构操作, 对其进行多角度、多层次的观察分析。系统最终实现了数据多维分析与查询, 为网站评估决策分析提供了有力支持, 对提高评估质量、促进评估发展具有现实意义。

**关键词:** 政府网站评估; 数据仓库; 联机分析处理

**中图分类号:** TP31

**文献标识码:** A

**文章编号:** 1673-629X(2011)10-0133-04

## Application of Data Warehouse and OLAP Technique in Government Website Appraisal

SHOU Zhi-qin, LIU Bo

(School of Management, Hefei University of Technology, Hefei 230009, China)

**Abstract:** The important component of the E-government construction is the government website appraisal, which has accumulated massive data. At present, the signification research subject of developing the government website assessment is how to use these precious information resource serves for management and the decision-making of the website appraisal. It introduced the data warehouse and the OLAP technology into the analysis system of the government website appraisal. Then, the data cube of government website evaluation is constructed and multi-angle and multi-level observation and analysis are made for it. As a result, the system has eventually realized the data multi-dimensional analysis and the inquiry, which has provided the powerful support for the decision-making of policymakers, furthermore, the system has the practical significance for enhancing the appraisal quality and improving appraisal development.

**Key words:** government website appraisal; data warehouse; online analytical processing

## 0 引 言

2006 年以来, 作者所在单位受安徽省政府网站建设办公室的委托, 作为第三方评估机构, 对全省省直厅局部门、地市级政府、县(区)级政府等 170 多家政府系列网站开展了连续五年的、较为系统的网站绩效评估工作<sup>[1]</sup>。随着评估工作的逐年推进, 积累了丰富的评估数据。由于评估数据电子化初期对数据标准存在一个不断摸索的过程, 大量数据存储 excel 中, 不能对这些数据进行集中存储查询。同时随着评估工作的不断深入, 评估目标越来越复杂, 评估难度越来越大, 由于缺乏集中存储和管理, 这些历史数据不能进行有效的统计、分析和评估。传统 MIS 仅仅局限于分类、汇总等简单的数据分析功能, 对数据的整合、利用还不

够充分, 不能满足对大量数据进行管理与分析的需求, 对决策支持的能力明显不足。而数据仓库具有的主题性、集成性、非易失性、时变性的特点, 为处理海量的、复杂结构的、持久的、共享的数据提供了技术可能; 将其与 OLAP 技术相结合, 能够为政府网站评估(以下简称网站评估)提供强有力的辅助决策支持。

## 1 网站评估需求特点

### 1.1 网站评估特点

通过对网站评估的业务需求进行分析, 总结出网站评估具有以下特点:

(1) 网站评估面向主题的多样性。安徽省政府网站评估指标体系按行政划分, 可分为三级——省直厅局级、地市级、县区级; 每级指标体系按功能划分, 可分为四大功能主题——信息公开、在线办事、公众参与、网站设计。

(2) 历史数据的稳定性<sup>[2]</sup>。网站评估数据反映了网站客观情况的历史真实性, 它是不同年限、不同地区

收稿日期: 2011-03-08; 修回日期: 2011-06-11

基金项目: 安徽省政府网站群年度绩效评估研究与分析(2006-2010)

作者简介: 寿志勤(1958-), 男, 浙江诸暨人, 副研究员, 硕士生导师, 研究方向为电子政务、政府网站评价理论与方法、信息化规划。

级别数据的集合,以及基于这些集合的统计、汇总和重组所产生的数据,这些数据一旦形成就不会发生变化。数据的稳定性是保证分析决策正确的基础条件。

(3)评估的时变性。网站评估工作逐年开展,评估目标逐年改进,其决策是与时间有关的。网站评估中的数据的改变是以年为单位进行阶段式改变而不是随机连续的改变,这种改变具有阶梯式表现形式。

### 1.2 建立网站评估数据仓库的必要性

作者所在单位开发过一个网站评估管理系统,此系统只是用来在线采集数据,并将数据存入关系数据库中。此系统主要进行操作性处理,通过简单关系连接的多个表读写少量的元组,其查询类型是一种联机事务处理(OLTP)模式。网站评估决策支持系统应该提供一个综合的、面向分析的环境。此时,基于数据库的决策支持系统显得力不从心,急需建立一个基于数据仓库(DW)的决策支持系统。因此,应在现有系统基础上进行迭代开发,建立网站评估数据仓库,现有系统作为数据仓库的数据来源。数据仓库的“多维表示”能够提供易于管理的信息视图。基于数据仓库的OLAP<sup>[3]</sup>“多维操作”(切片、切块、旋转、钻探)能够满足网站评估主题多样性的需要。基于数据仓库的网站评估决策支持系统能够提供灵活的查询分析功能,能对网站评估详细信息进行汇总,提供对决策有用的重要数据。

## 2 网站评估数据仓库体系结构

数据仓库概念的形成是以 W. H. Inmon 于 1990 年出版的著作“Building the Data Warehouse”(中译名为《数据仓库》)为标志,他在该书中给出了数据仓库的定义:“数据仓库是面向主题的、集成的、时变的以及非易失的数据集合体”<sup>[4]</sup>。数据仓库是在关系数据库基础上发展起来的,主要用于数据清理、整合、在线分析以及数据挖掘。数据仓库是建立在事务处理数据源系统之上的集成和分析处理系统,是管理信息系统的“上层建筑”。数据仓库本身的结构是由多层叠加的系统构成的,网站评估数据仓库的金字塔结构和层次

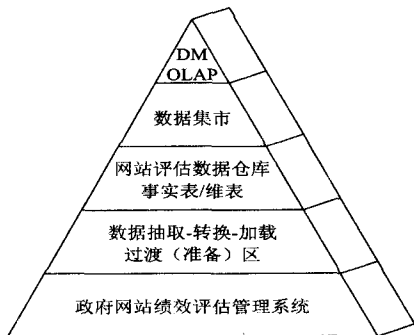


图 1 网站评估数据仓库的金字塔结构和层次

如图 1 所示。

从图 1 可知,网站评估数据仓库分为 5 个层次。

(1)数据源层:网站评估管理系统作为数据来源。

(2)数据后端处理层:数据源中数据经抽取、转换、加载(Extraction、Transformation、Loading, ETL)最终成为网站评估数据仓库所需的数据<sup>[5,6]</sup>。

(3)数据仓库及其管理层:此层包括数据仓库和数据仓库管理两个部分。数据仓库负责存储用于决策分析的数据,即评估事实表和维表中的数据。数据仓库管理则负责管理数据仓库,包括数据模式定义、数据操纵、数据控制、元数据的管理等。

(4)数据集市层:此层为局部数据仓库,为指定的应用提供数据支撑。数据集市建立在网站评估数据仓库之上。网站评估数据集市是针对特定应用需要的多粒度的综合数据,数据集市的主题是数据仓库中的局部性主题,根据指标体系结构组织数据。

(5)数据仓库应用层:数据仓库是为应用服务的,数据仓库应用层从网站评估数据集中提取数据,针对特定主题进行联机分析处理和数据挖掘等。

## 3 网站评估系统数据仓库设计

### 3.1 主题确定

网站评估数据仓库注重的是查询而不是事务,设计的目标是让决策者能够根据不同的目标 and 需求访问数据,而确定主题是组织数据仓库中数据的前提。网站评估数据仓库主要按照“评估”这一主题来组织数据。按指标体系结构,可将主题进行细分:每级指标可确定若干评估子主题,主题下可包含若干下一级主题,每个子主题需要一个数据集合体作支撑<sup>[7]</sup>。如一级指标可划分为信息公开、在线办事、公众参与、网站设计四个主题。按指标体系结构划分数据仓库相关主题如图 2 所示。

一级指标 (一级主题)	二级指标 (二级主题)	三级指标 (三级主题)	指标细则
信息公开	主动公开信息	概况信息	对部门职责信息的公开情况 3 对领导信息的公开情况 3 对部门机构设置、司局职责的公开情况 4
		法规文件	.....
		.....	.....
		.....	.....
在线办事	服务框架	服务框架	结合部门职能、用户对象设计服务框架 3 服务框架是否能全面覆盖部门职能 4 能否以事件为核心整合各项服务资源 3
		.....	.....
公众参与	咨询投诉	渠道功能	投诉咨询渠道种类 2 提供使用帮助或指南 2 能按主题提供常见问题列表 4 能对信件答复进行满意度投票 2
		答复情况	.....
		.....	.....
		.....	.....
网站设计	网站响应	首页响应时间	主页或者网页内容 5
		关键搜索模块的响应时间	5

图 2 按指标体系结构划分主题

### 3.2 概念模型设计

设计中采用一种信息包图的方法在平面上展开立方体。信息包图是一种直观的、紧凑的概念模型设计工具,它用二维表格的形式来反映用户使用数据仓库进行多维处理的需求<sup>[8]</sup>。信息包图由事实、维度、类别组成,网站评估概念模型的信息包图如图 3 所示。

概念模型的设计步骤如下:

1) 确定度量。度量是网站评估的量化指标,是评估的事实依据,决策者根据度量数值对评估结果进行统计分析。网站评估系统的度量指标为网站目标总体得分、单个网站得分、各级指标分项得分等。

2) 确定维度。维度是事实的属性信息,也是考察事实的角度。网站评估系统主要从网站维、时间维、指标维三个角度对评估事实进行考察<sup>[9]</sup>。

3) 确定粒度。粒度体现了维度的层次结构,体现了数据的详细程度。网站评估系统将网站维根据行政级别的不同,划分为三个层次:省、市、区县;指标维根据指标体系结构,划分为四个层次:一级指标、二级指标、三级指标、指标细则。

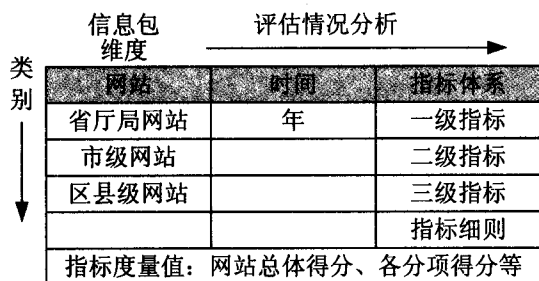


图 3 网站评估分析信息包

### 3.3 逻辑模型设计

网站评估展现多维数据结构的雪花模型<sup>[9]</sup>是以“信息包图”为基础,同样可分为度量、维度、粒度三部分(如图 4 所示)。度量以事实表表示,放在雪花模型的中央,它存放多维表中的主要事实;维度以维表表示,放在雪花模型的四周,它用来维护多维结构中的维值;粒度体现维中的多个层次。“维”中维层次的一个取值称为维成员,当每个“维”确定一个取值(即确定维成员)时即可获得多维结构中的一个确定的数据量(称单元)。多维结构的这种组成方式可用:(维 1, 维 2, ..., 维  $n$ , 单元)表示,称多维数组,在这个多维数组中对每个“维”确定一个维成员后即可唯一确定单元。如图 4 所示的多维数据结构,它的维成员取值可以是:1、时间维成员:2010;2、网站维成员:ahedu;3、指标维成员:tag0001。根据这 3 个取值可以在多维结构中获得单元值:2.4,即(2010, ahedu, tag0001, 2.4)。它反映了 2010 年安徽省教育厅网站的某项具体指标细则 tag0001 的评分为 2.4 分。通过对事实表进行各种统

计(包括求和、求平均值、总计等)可以求得不同粒度的单元值。

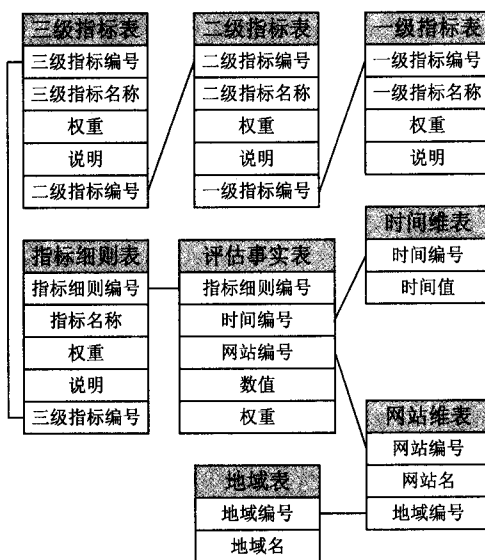


图 4 雪花逻辑模型

## 4 网站评估数据仓库的 OLAP

OLAP 与数据仓库有着紧密的关系。数据仓库的建立解决了依据主题进行数据存储的问题,而 OLAP 在某种程度上解决了如何利用这些存储数据的问题,OLAP 可能是在数据仓库中使用信息的主要方式。通常,数据仓库面向分析主题进行设计,并针对各个主题合理规划 OLAP 的维度和粒度,这样数据仓库才能更好地支持 OLAP。OLAP 技术的多维结构操作是核心所在,这些操作包括切片、切块、旋转、上卷及下钻(见图 5),各种分析动作以剖析数据为目标,使分析人员能多角度、多侧面、多层次地观察多维结构中的数据。

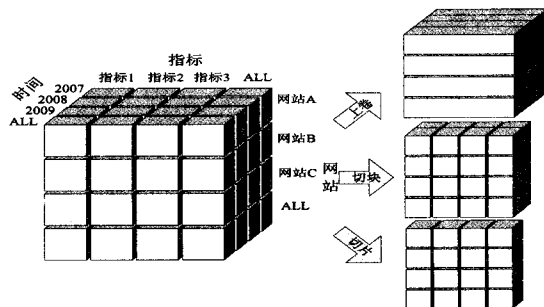


图 5 数据立方体钻探、切块、切片

### 4.1 钻探

钻探操作是用户对数据深度的操作,在多层数据中能通过钻探操作使用户能自由往返于不同深度数据层次中,钻探操作包括上卷(roll-up)和下钻(drill-down)。

上卷操作增加数据聚合程度,并从层次结构中清除细节层次。细节数据层的事实表存在大量的数据,不利于查询,也不利于为决策支持服务,这时就需要上

卷操作。上卷操作用来汇总细节数据,在指标维上进行上卷操作,可以分别汇总各个指标层次的数据,分别得到三级指标汇总数据、二级指标汇总数据、一级指标汇总数据,同时可对这些数据求和、求平均值、求最大值、求最小值。

下钻操作是上卷操作的互补操作,它可以降低数据集程度,并向层次结构添加新的细节层次。通过上卷和下钻操作,能够得到各个维度上的不同层级的数据。

#### 4.2 切片(slice)和切块(dice)

切片是把某个维度设为定值,从而减少立方体维度数量的操作<sup>[10]</sup>。网站评估中,每年会导入新的数据,当需要对最新数据进行分析时,就需要在时间维上进行切片。例如要对2010年的数据进行分析,就需要在时间维上把时间设定为2010年。这样可以得到由指标和网站组成的二维表格。同理,需要对某个网站的历史指标数据进行分析时,可将网站设为一个定值。

切块是限制多维数据的某一维的取值区间,当这一区间只取一个维成员时,即得到一个切片<sup>[11]</sup>。如果要对政府网站在2009年和2010年的评估数据进行对比,这时就需要进行切块操作,时间维上设定为2009-2010年。

#### 4.3 旋转(rotate)

根据多维比喻法,旋转数据就是旋转立方体,从而可以从一个新视角入手重新安排立方体单元。通过交换网站、指标横纵坐标位置,可以得到一张新的表格数据,从而可从不同维度比较分析数据。

### 5 基于数据仓库和 OLAP 的网站评估分析系统的实现

本系统采用 Microsoft 的 SQL Server 2005 系列的数据库管理工具作为数据仓库的组建工具,采用 SQL Server 2005 Analysis Services (SSAS) 构建数据立方体(Cube)进行 OLAP 分析。OLAP 的分析结果展示可通过 SQL Server 2005 Reporting Services (SSRS) 工具实现,根据不同主题显示相应的图表和报表。同时本系统支持基于 Web 的数据分析和展示功能,能为多用户提供在线服务。

系统提供 B/S 的数据展现方式,使用 J2EE 架构实现,选用 Tomcat 作为 Web 服务器。系统采用 MVC (Model-View-Controller),即“模型-视图-控制器”软件设计模式,系统的服务请求及响应过程实现为:用户从客户端通过 HTTP 链接向 Web 服务器发送请求;控制层 Servlet 接收用户从控制界面发送的命令,并调用

Model 层的 JavaBean 进行业务逻辑处理<sup>[12]</sup>。所有业务逻辑和数据处理都封装在不同的 JavaBean 当中,JavaBean 中封装了对数据进行切片、切块、钻探的 SQL 语句。同时,初始化连接池和建立、管理、关闭连接等操作也封装在 JavaBean (ConnectionPool 类)中;返回的主题数据传入 JFreeChart 的 API 生成相应的辅助分析直方图,在输出结果的 JSP 页面中动态地设定这些图的连接。

### 6 结束语

文中详细介绍了如何将数据仓库技术与网站评估进行有机的结合,对评估数据进行了有效的存储与管理,为 OLAP 技术进行多角度、多层次地统计和分析数据提供了基础。网站评估系统的开发,更新和发展了传统的评估方法与手段,改善了评估的效果和效率。同时,随着评估工作的不断深入、需求的不断增加,需要对系统进行不断的扩展与完善,实现系统的迭代开发,建立更为强大和科学的决策支持系统。

#### 参考文献:

- [1] 寿志勤,陈鑫.安徽实验:第三方政府网站绩效评估机制研究[J].中国科技论坛,2010(2):114-119.
- [2] 王闯,余本功.数据仓库与 OLAP 技术在科技部门网站评估中的应用[J].合肥工业大学学报(自然科学版),2007,30(12):1661-1664.
- [3] Codd E F, Codd S, Salley C. Providing OLAP (on-line analytical processing) to User-analysts: an IT mandate [R]. [s. l.]: IBM Research Lab, 1933.
- [4] 于红蕾,华庆一,刘燕玲,等.数据仓库在电信统计分析中的应用[J].计算机技术与发展,2007,17(8):201-204.
- [5] Inmon W H. Building the data warehouse [M]. 2nd ed. [s. l.]: John Wiley, 1996: 50-93.
- [6] Simitsis A, Vassiliadis P, Sellis T. Optimizing ETL Processes in Data Warehouse [C]//ICDE. Tokyo, Japan: [s. n.], 2005: 564-575.
- [7] 杨国梁,关忠诚,石兵.数据仓库技术在中国科学院评估数据分析中的应用[J].科研管理,2007,28:114-120.
- [8] 杨兴凯.基于本体的政务数据仓库构建方法研究[J].计算机工程与设计,2010,31(7):1492-1494.
- [9] 刘黎志.基于数据仓库的指标体系研究[J].计算机技术与发展,2007,17(1):196-198.
- [10] Kimball R. The Data Warehouse Lifecycle Toolkit [M]. 2nd ed. New York: John Wiley, 2008.
- [11] 徐洁磐.数据仓库与决策支持系统[M].北京:科学出版社,2005.
- [12] 刘浪,王丽亚,黄海量.基于 Web 的数据仓库解决方案[J].计算机工程,2005,31(1):92-94.