第16卷 第2期
2006年2月

计算机技术与发展
COMPUTER TECHNOLOGY AND DEVELOPMENT

Vol.16 No.2
Feb. 2006

主题地图在关系数据库中的应用研究

吕云屏,吴 洁
(南京航空航天大学 计算机应用研究所,江苏 南京 210016)

摘 要:作为ISO确立的新标准,主题地图用来描述知识结构及其内在关联。它提供了语义级的数据导航和组织方式,是一个表达和交换结构化信息的元数据模型。文中在介绍了主题地图的相关概念后,设计了一种基于主题地图、扩展关系数据库的方法,并利用主题地图的描述语言XTM导出数据库,为多个关系数据库之间的数据交换提供了新的理念。

关键词:主题地图;关系数据库;XTM
中图分类号:TP311.132.3 **文献标识码**:A **文章编号**:1005-3751(2006)02-0087-03

Application and Research of Topic Maps in Relational Database

LÜ Yun-ping, WU Jie
(Institute of Computer Application, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Topic maps are a new ISO standard for describing knowledge structures and associating them with information resources. As a data model of describing and exchanging information, they provide a way of navigating and organizing semantic data. In the paper, describe the related concept of topic maps at first. Then a method of how to extend relational database using topic maps is suggested. At last, using XTM educe database that is the description language of topic maps in order to provide the new idea for exchanging data in multi-database.

Key words: topic maps; relational database; XTM

0 引 言

主题地图(Topic Maps)是ISO提出的新标准,在ISO/IEC 13250中对主题地图有如下的描述[1]:"主题地图定义了一个多维的主题空间——在这个空间里,各个位置表示不同的主题。根据从一个主题到另外一个主题所需要转换主题的次数,就可以计算出任意两个主题间的距离,而它们之间的转换关系定义了从一个主题到另外一个主题的路径,路径则用转换的主题来表示。"主题地图的出现为数据的搜索、组织和分析提供了一种有效的方式。它提出了一种基于主题的元数据组织和描述方式,提供了语义级的数据导航和组织方式,是一个表达和交换结构化信息的元数据模型[2]。

20世纪80年代发展起来的关系数据库是目前使用非常广泛的一种数据管理系统。这种数据库具有数据结构化、最低冗余度、较高的程序与数据独立性、易于编制应用程序等优点。这种传统的数据库可以在已有的表结构上很好地维护数据,然而当用户希望打破现有的表结构,向数据库中存储进新的数据信息,对数据库进行扩展时,就需要重新设计数据库的架构。对于大型数据库,这种更新无疑代价巨大,因此对数据库的设计者提出了很高的要求,希望他们最初在设计数据库时考虑完善周全,有预见性。尽管如此,数据库的更新扩展仍然无法避免。

主题地图的出现为上述问题的解决提供了新的思路。通过主题地图,可以灵活有效地实时扩展关系数据库,并实现不同关系数据库之间的数据交换。

1 主题地图的数据模型

主题地图基本模型由TAO三要素构成[3,4],即主题(Topic)、联系(Association)和事件(Occurrence)。

* T——主题(Topic):

主题是主题地图中知识的基本单元,它可以是任何东西,包括人、实体、概念等。任何不存在或不具有具体特征的事物也可以作为主题。主题可以划分为不同的类别,称为主题类型(Topic types)。主题类型就是主题所归属的类别,一个主题可以归属到一个以上的主题类型,主题类型在主题地图中也被认定为一个主题。

* A——联系(Association):

主题之间可以用联系来显示其语义关系,这样的主题通常称为成员。联系的形式可以是一对一、一对多或者多对多。联系也有联系类型(Association types),联系类型把具有相同关系的主题汇集成类,这样的主题(常为主题类型)通常被称为角色。联系类型有助于增强主题地图的表达能力。联系是主题地图的主要功能。因为,将存在于主

收稿日期:2005-05-12
作者简介:吕云屏(1981—),女,江苏南京人,硕士研究生,研究方向系统集成与企业信息化;吴 洁,副教授,主要从事计算机教学和研究工作。

题之间的各类关系,透过联系的组织与联结后,将形成某一领域知识的知识网络。

* O——事件(Occurrence):

一个主题可以联结至一个或者多个在某种层面上被视为与该主题相关的信息资源,这样的信息资源称为该主题的事件。在实际使用的过程中,事件通常是指储存于全球信息网里的任何形式的资源,即可以由 URI 存取到的资源。

主题、主题类型和事件的概念使得人们能够根据主题来组织信息资源并建立索引;联系和联系类型的概念则使得人们能够描述主题间的关系,建立语义关系网络,这样就形成了最基本的主题地图。当然,主题地图所包含的基本概念远远不止这些,但是对于简单的应用来说已经足够了。

假设有一个部门 - 员工的的数据库,包含 department 和 staff 两个表,用来描述部门和员工的信息,如下所示:

(1) department(Department_id varchar(18), Manager_id varchar(18), Department_name varchar(20))

Department_id	Manager_id	Department_name
D_001	S_001	市场部

(2) staff(Staff_id varchar(18), Department_id varchar(18), Staff_name varchar(20), Birthday datetime(8), Salary int(4))

Staff_id	Department_id	Staff_name	Birthday	Salary
S_001	D_001	张扬	1980-4-5	10000

对表中提供的信息,可以按照 TAO 三要素的概念进行分析。主题包括:部门,员工,市场部,张扬。其中"部门"和"员工"是主题类型;联系包括:张扬任市场部经理,其中"任职"为联系类型;事件包括:出生日期,薪水(这也可以理解为主题)。

2 主题地图的描述语言 XTM

XTM 是 XML Topic Maps 的缩写,它是基于主题地图规范的描述语言,采用 XML 的语法结构。XTM 主要定义了用于描述主题地图的 DTD 文件,提供了描述结构化信息的语法和模型,该语法可定义主题、主题与主题间的联系等。

XTM 中将使用到这样一些元素[5]:<topicRef>,<subjectIndicatorRef>,<scope>,<instanceOf>,<topicMap>,<topic>,<subjectIdentity>,<baseName>,<baseNameString>,<variant>,<variantName>,<parameters>,<association>,<member>,<roleSpec>,<occurrence>,<resourceRef>,<resourceData>,<mergeMap>。

对于上面的例子,XTM 的描述如下(部分):

```
……
<topic id="staff"><baseName><baseNameString>Staff</baseNameString></baseName></topic>
<topic id="department">…</topic>
<topic id="assignment">…</topic>
<topic id="birthday">…</topic>
<topic id="salary">…</topic>
<topic xmlns=""id="S_001">
  <instanceOf><topicRef xlink:href="#staff"/></instanceOf>
  <baseName><baseNameString>张扬</baseNameString></baseName>
  <occurrence>
    <instanceOf><topicRef xlink:href="#birthday"/></instanceOf>
    <resourceData>1980-4-5</resourceData>
  </occurrence>
  <occurrence>
    <instanceOf><topicRef xlink:href="#salary"/></instanceOf>
    <resourceData>10000</resourceData>
  </occurrence>
</topic>
<topic xmlns=""id="D_001">…</topic>
<association xmlns="">
  <instanceOf><topicRef xlink:href="#assignment"/></instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#department"/></roleSpec>
    <topicRef xlink:href="#D_001"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#employee"/></roleSpec>
    <topicRef xlink:href="#S_001"/>
  </member>
</association>
……
```

3 主题地图对关系数据库的扩展

利用主题地图的概念,可以在保持已有数据库架构的基础上实时扩展,再将更新了的关系数据库导出成为 XTM 文件,实现多个关系数据库之间的数据交换。数据库的扩展,分为基本数据类型的扩展和自定义数据类型的扩展。前者指的是新加入的信息内容用数据库中已经定义了的基本数据类型即可以描述;后者是指所需添加的信息并不能用 int,char 等现有简单数据类型定义,因而称它为用户自定义类型数据信息。对于这两种不同类型的扩展,采用不同的方法处理。

以上研究,在扬州电子有限公司的信息系统扩建项目中得到了很好的应用。以其某部门的员工数据库为例,为阐述方便,将它简化为上述部门 - 员工数据库,假设现需

要给员工添加年龄和籍贯信息,其中籍贯信息只能从如下几个省份中选取:江苏、浙江、陕西、甘肃。

根据上面的假设,分析出员工"年龄"可以用常用的int型数据表示;而"籍贯"并不能简单的用char或varchar等字符类型描述,因为该信息的取值有一定的制约条件,所以它属于自定义数据类型。

3.1 基本数据类型扩展

这种方式的扩展较简单。在数据库中新建事件信息表occurrence_type和occurrence,如下所示:

(1)occurrence_type(occurrence_type_id varchar(18), name varchar(20),datatype varchar(10))

(2) occurrence(occurrence_id varchar(18),instanceOf varchar(18),topic_id varchar(18),resource varchar(50))

向表中插入数据后,如表1、表2所示。

表1　occurrence_type

occurrence_type_id	name	datatype
Occurrence_type_1	Age	int

表2　occurrence

occurrence_id	instanceOf	topic_id	resource
Occurrence_1	Age	S_001	25

3.2 自定义数据类型扩展

与基本数据类型扩展比较,这种方式的扩展相对复杂,需要建立主题信息表用来描述该主题内容,建立联系信息表用来表示新增主题与原主题之间的联系。

●新建数据表topic_type和topic,如下所示:

(1)topic_type(topic_type_id varchar(18),name varchar(20))

(2) topic(topic_id varchar(18), instanceOf varchar(18),name varchar(20))

向表中插入数据后,如表3、表4所示。

表3　topic_type

topic_type_id	name
Topic_type_1	Province

表4　topic

topic_id	instanceOf	name
Topic_1	Province	江苏省
Topic_2	Province	浙江省
Topic_3	Province	陕西省
Topic_4	Province	甘肃省

●新建数据表association_type和association,如下所示:

(1) association_type (association_type_id varchar(18),name varchar(20),role_type_1 varchar(20),role_type_2 varchar(20))

(2) association(association_id varchar(18),instanceOf varchar(18),member_1 varchar(20),member_2 varchar(20))

向表中插入数据后,如表5,表6所示:

表5　association_type

association_type_id	name	role_type_1	role_type_2
Association_type_1	Region	Staff	Province

表6　association

association_id	instanceOf	member_1	member_2
Association_1	Region	张扬	陕西省

3.3 数据库的导出

为了能够使得不同的关系数据库间进行数据交换,可以将数据库导出成为XTM文件,因为XTM文件遵循XML语法,所以即可以使用处理XML的各种方法对生成的文件进行分析,实现数据库之间的数据交换。

首先,对原有数据库架构进行剖析。根据TAO三要素的含义,识别出数据信息中的主题、联系和事件,如小节1中所示;将这些信息分别用XTM规定的元素节点进行描述,生成XTM文档,如小节2中所示。

接下来,处理新加入数据库中的数据信息。分为六步:

第一步:为所有主题类型生成节点<topic>,包括其属性id和子节点<baseName>(包含属性<baseNameString>)。如下SQL语句选择出主题类型。

Select topic_type_id as id, name from topic_type

第二步:为所有联系类型生成节点<topic>,包括其属性id和子节点<baseName>(包含属性<baseNameString>)。如下SQL语句选择出联系类型。

Select association_type_id as id, name from association_type

第三步:为所有角色生成节点<topic>,包括其属性id和子节点<baseName>(包含属性<baseNameString>)。这里的角色可能会和主题类型有所重复,因此只需要添加一次即可。为了区分出<topic>的属性id和<topic>的子元素<baseName>,采用如下SQL语句选择角色。

Select distinct'role_type_'&role_type_1 as id, role_type_1 as namefrom association_type

Select distinct 'role_type_' &role_type_2 as id, role_type_2as name from association_type

第四步:为所有事件类型生成节点<topic>,包括其属性id和子节点<baseName>(包含属性<baseNameString>)。如下SQL语句选择出事件类型。

Select occurrence_type_id as id, name from occurrence_type

第五步:为所有常规的主题生成<topic>,包括其属性id和子节点<baseName>(包含属性<baseNameString>)。如下SQL语句选择出主题。

Select topic.id&'_'&topic_type.name as id, topic_type.topic_type_id as topic_type, topic.name from topic_type, topic where topic_type.name=topic.instanceOf

第六步:为数据库中的所有联系生成<association>节点,包括其属性id和子节点<baseName>(包含属性<basedName String>)。如下SQL语句选择出联系。

(下转第178页)

表 1 用不同训练样本训练后进行拟合的 R 值的比较

	训练样本取法	多项式核函数	rbf 核函数	混合核函数
HDL	隔值取法 1	0.903	0.737	0.92
	隔值取法 2	0.886	0.719	0.908
	前三分之一	0.883	0.661	0.893
LDL	隔值取法 1	0.886	0.703	0.898
	隔值取法 2	0.823	0.703	0.879
	前三分之一	0.816	0.613	0.859
VLDL	隔值取法 1	0.51	0.594	0.685
	隔值取法 2	0.532	0.547	0.59
	前三分之一	0.437	0.589	0.644

表 2 不同 σ 参数值下进行回归的 R 值的比较

σ	HDL	LDL	VLDL
0.01	0.92	0.898	0.685
0.1	0.92	0.898	0.683
0.2	0.923	0.898	0.675
0.5	0.929	0.886	0.601
1	0.895	0.856	0.435
2	0.874	0.859	0.26
5	0.57	0.582	0.0861

表 3 不同的 C 值下平均训练时间的比较

C 值	∞	10^7	10^5	10^3	10^2
平均训练时间	11.1	3.6	2.9	2.5	2.4

不仅如此，实验中还发现，C 值的选取影响训练时间的长短，C 值越小平均训练时间越短，但当 C 值过小（$C \leqslant 10^2$）时，实验结果的精度会下降。表 3 是 C 取不同值时平均训练时间的值，因此本实验中采用 $C = 1000$ 是较合理的，并且有效提高了 SVM 方法的训练速度。以上实验均利用 Matlab 6.1 编程，运行于 Pentium Ⅳ/2G，256M 内存 PC。

3 结束语

文中简要介绍了由混合核函数构造的支持向量机，并将其运用于函数拟合中。通过对 3 种不同类别血浆脂蛋白样本与其血浆胆固醇的含量的测定，验证了选择这种混合核函数的实验具有很好的效果，实验中 VLDL 的精确度有明显提高，而且本实验中训练时间只有 2.5 秒左右，很好地解决了训练速度慢的问题。当然，还可以考虑用其它的核函数来进行混合，形成不同的混合核函数，譬如将两个或多个 RBF 核函数混合，或者将两个或多个多项式核函数混合形成混合核函数，或许会得到更好的效果，以便找到选择最优核函数的某些规律，这是以后值得研究的一个课题。

参考文献：

[1] Smits G F, Jordaan E M. Improved SVM Regression using Mixtures of Kernels[A]. Proceedings of the 2002 International Joint Conference on Neural Networks[C]. Hawaii: IEEE, 2002.2785 - 2790.

[2] 边肇祺，张学工. 模式识别[M]. 北京：清华大学出版社，1999.

[3] Zhang Li, Zhou Weida, Jiao Licheng. Wavelet Support Vector Machine[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 2004, 34(1): 34 - 39.

[4] Zhang Sheng, Liu Jian, Tian Jin - wen. An SVM - based Small Target Segmentationand Clustering Approach[A]. Proceedings of the Third International Conference on Machine Learning and Cybernetics[C]. Shanghai: IEEE, 2004. 3318 - 3323.

[5] 丁 蕾，陶 亮. 支持向量机在胆固醇测定中的应用[J]. 安徽大学学报(自然科学版)，2005(2): 60 - 63.

(上接第 89 页)

Select association_type. association_type_id, 'role_type_' & association_type. role_type_1 as role_1, associaton. member_1 as member_1, 'role_type_' &associaton_type. role_type_2 as role_2, association. member_2 as member_2 from association_type, association where association_type. name = association. instanceOf

4 结束语

主题地图是一个新兴的 ISO 标准，它提供了一种用于组织信息的系统。文中首先提出了主题地图的概念，描述了主题地图的 3 要素：主题、联系和事件的含义，并介绍了其描述语言 XTM。最后，将主题地图的概念方法应用到扬州宝军电子有限公司的信息系统扩建项目中，对其现有数据库，设计了一种扩展关系数据库，实现多个关系数据库之间数据交换的方法。

随着人们对主题地图的不断认识了解，相信这样一种数据描述与组织的技术，必将在更加广泛的领域中得以应用。

参考文献：

[1] 秦铁辉，郭延吉，孙 琳. 信息时代的"全球定位系统"——主题地图[J]. 江西图书馆学刊，2005, 35(1): 1 - 3.

[2] 张佩云，吴 江，贾 晖. 主题地图标准及其应用研究[J]. 安徽大学学报(自然科学版)，2004, 28(3): 19 - 22.

[3] Pepper S. Chief Strategy Officer The TAO of Topic Maps [EB/OL]. http://www. ontopia. net/topicmaps/materials/tao. html, 2002.

[4] Garshol L M. What Are Topic Maps? [EB/OL]. http://www. xml. com/pub/a/2002/09/11/topicmaps. html, 2002.

[5] TopicMap. Org XML Topic Maps (XTM) 1.0. [EB/OL] http://www. topicmaps. org/xtm/index. html, 2000.