

一个基于本体论的职工能力查询系统

廖明宏

(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)

摘要 在竞争日益激烈的市场经济中,企业界越来越清楚地认识到存在于企业里的各种知识、信息是一笔非常重要的财富。一门新兴的交叉学科—企业知识管理得到越来越多学者的关注。即采用先进的信息技术,形式管理知识资源,以便在企业范围内访问和使用知识。本体论作为一种新的知识表示方法成为企业知识管理系统不可缺少的组成部分。本文介绍的企业职工能力查询系统是企业知识管理系统的一个重要组成部分,其基于本体论的查询思想比传统的基于关键字的查询思想更接近人们的思维习惯,其查询结果更能满足用户的要求。

关键词 企业知识管理 本体论 启发式查询

1 引言

在当今的市场经济中,企业面临着日益激烈的竞争环境。在企业减员增效的同时,企业也丢失了员工的重要知识。与此同时,随着人才市场的日益火爆,人才的流动也将带走重要的知识信息。随着市场的国际化,地区间的差异也影响了知识的交流与有效访问。因此,一门新兴的交叉学科:企业知识管理也应应运而生^[1]。企业知识管理主要研究如何形式化地管理知识资源,尤其是采用先进的信息技术,以便在企业范围内有效地访问和使用知识。企业知识管理涉及到知识的产生、表示和查询等几个主要研究领域。在知识表示方法上,本体论(Ontology)被认为是企业知识管理重要的组成部分。本体论就是对概念化的精确描述^[2]。这个概念最初来源于哲学界,它是研究存在的本质的哲学问题。然而近年来,这个词被应用到计算机界,并在人工智能、计算机语言学以及数据库理论中扮演着越来越重要的角色^[3]。本体论的主要目的是精确地表示那些隐含的信息,使得它们可被软件系统重用和共享。在企业知识管理系统中,本体论描述涉及到任务的分类学,这些任务定义了用于系统的知识,本体论定义了用于知识管理系统的共享字典,它有利于系统的通信、查找、存储和表示。开发和维护企业范围内的本体论需要不断地努力,并随着时间的推移而不断完善。

作为企业知识管理系统一个重要的组成部分,企业职工能力查询系统(CRS)用于管理职工的各种知识能力,并为用户提供信息支持。该系统的应用领域很广,最典型的一个应用就是工作部门人员配置。根据工作部门对员工的特殊要求,利用能力查询系统,找到合适的人选。另一个应用就是商品的售后服务。当商品出现问题需要找专家解决时,能力查询系统能找到合适的专家(这一点在一个大型企业中尤为重要)。

CRS不同于一般的数据库系统,它是在本体论的支持下建立起来的。传统的数据库系统主要是基于关键字查询,有时其查询结果是不能令人满意的。比如查找“知识表示”这个关键字没有结果,并不等于没有人懂得“知识表示”。事实上可能有人懂得“产生式系统”,它是一种知识表示方法,按常理,他应是懂得“知识表示”的人选。当然在传统数据库中我们可以通过输入“产生式系统”找到合适人选。但我们不能要求使用者知道“产生式系统”是“知识表示”的一种。借助于本体论,我们可以使系统本身在查询时考虑到这些因素,因此其查询的结果更能满足用户的需求(想进一步了解,参见参考文献4)。

本文主要介绍一个企业职工能力查询系统的设计,其中包括系统的体系结构、企业本体论模型和基于本体论的信息检索方法,最后还将总结初步的研究成果。

2 企业职工能力查询系统的体系结构

在现代化企业中,人们总是建立Intranet作为其基本的信息处理设施。Intranet是由一组附属的客户机连接而成的网络,它们使用标准的Internet协议,尤其是TCP/IP和HTTP协议。为安全起见,它与外界之间总是隔着一道或几道防火墙。为使我们设计的能力查询系统可在Intranet上运行,我们采用客户—服务器模型,并用Java语言编程。能力查询系统的体系结构见图1。

在客户端,用户可输入姓名、项目或直接输入能力,其中每个域还支持复合关系表达式,如:(数据库AND逻辑)OR演绎数据库。输入界面用Java的小应用程序(applet)编写,因此可被网络浏览器如Netscape或Internet Explorer调用。输出界面是一张二维表,它记录了职工的基本信息以及相关的能力信息。该二维表是由服务器自动建立的,并以HTML文件格式存放,因此在客户端通过访问这个文件就可在终端上显示查询结果。

收稿日期:1999年6月10日

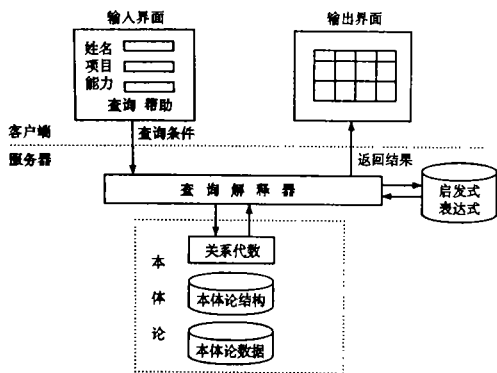


图1 能力查询系统的体系结构

系统的核心部分是放在服务器一端，其中包括查询解释器、启发式表达式和本体论等部分。查询解释器接收用户提交的查询条件，并与启发式表达式相结合，形成本体论的操作指令，查找本体论得到查询结果，并构造二维表返回给客户端。启发式表达式是一组基于本体论的，用于启发信息检索过程的公式。它是该系统区别于传统查询系统所在，因此，在后面的章节中还将对其作进一步的论述。本体论由本体论数据、本体论结构以及关系代数组成。本体论数据记载着企业、职工以及领域的有关数据及其相互联系的信息；本体论结构描述各个概念的数据结构。本体论的结构和数据以关系数据库形式存放在服务器上，Java可通过JDBC^[5]接口访问关系数据库。关系代数以对象为中心的，它提供一种对当前数据具有一定演绎能力的面向对象的访问方法。它为本体论提供基本的操作接口。

客户与服务器之间的通信是采用套接字方式(Socket)。对客户端的每一个查询请求，服务器产生一个服务线程(thread)来实现相应的请求，因此它可同时满足多个用户的查询要求。

3 企业本体论模型

任何一个信息或知识项都可由信息元模型、信息内容和信息的应用背景三方面加以描述^[6]。从企业本体论的角度看，它们分别对应于信息本体论、领域本体论和企业本体论(见图2)。



图2 企业本体论模型

其中，信息本体论用于描述信息元模型的字典，它描述了不同的信息源以及其相应的结构、访问和格式特性。领域本体论用于为信息源建模，主要反映概念间的不同关系。企业本体论用于描述信息的应用背景，主要刻划企业的组织机构以及处理模型。

这里，本体论是采用面向对象的方式描述的。一

个类可声明如下：

```
class (attribute1 : type1, ..., attributen : typen)
```

如果类1由类2派生来的，可用如下描述：

```
class1 : class2 (attributen+1 : typen+1, ..., attributen+m : typen+m)
```

在本体论中，类的属性建立一个语义网络，为能在语义网上的边标注对象，我们引入一个复杂的类型结构，即标注。一个标注属性可描述为：

```
attribute : class1 / class2
```

或(对于属性集)：

```
attribute : {class1} / class2
```

其中，class₂表示一个标注类。

一个类的对象可表示为：

```
class(attribute1=value1 / ann1, ...)
```

或：

```
name : class(attribute1=value1 / ann1, ...)
```

或(对于属性集)：

```
class(attribute1={value1 / ann1, ...}, ...)
```

现以领域本体论为例，说明本体论的数据结构及其具体的对象实例的表示形式。在领域本体论中，类content被定义为一个起点，其它类都由此类派生：

```
content() //抽象的超类
ann() //标注超类
concept: content(
    name: string,
    ld : {ld} //ld表示语言描述
    link: {concept}/conceptLink
)
conceptLink: ann( //由标注类ann派生
    name: string
)
subLink: conceptLink()
isco: subLink() //表示子概念
ispo: subLink() //表示子部分
ld(
    lang: language,
    name: string,
    abbr: string //缩写词
)
language(
    name: string
)
field: concept() //知识领域
eng: language(name="English")
ger: language(name="German")
sub: isco(
    name="is subconcept of"
)
isfield: isco(
    name="is subfield of"
```

领域本体论刻划了领域中概念及概念间的关系，因此，它的内容相当丰富。这里仅简单介绍计算机科学领域的一小部分描述。

```

)
cs: field(
  name="computer science",
  ld={csLde, csLdg},
Links={science/sub}
)
csLde: ld(
  lang=eng,
  name="computer science"
)
csLdg: ld(
  lang=ger,
  name="Informatik"
)
ai: field(
  name="artificial intelligence",
  ld={aiLde, aiLdg},
  links={cs/isfield}
)
aiLde: ld(
  lang=eng,
  name="artificial intelligence"
)
aiLdg: ld(
  lang=ger,
  name="kuenstliche Intelligenz"
)
.....

```

由本体论的内容描述可以看出，它与一般的数据库不同：一方面每个概念包含有语言特性即英语单词和德语单词两种形式；另一方面每个概念带有Link域，用以描述它与其它概念之间的关系。因此，从形式上看，本体论的内容本身就构成一个语义网络。

4 基于本体论的查询算法

我们设计的能力查询系统与传统的数据库系统的区别主要体现在查询算法上。众所周知，数据库查询是依赖于关键字匹配，若匹配失败返回空记录，不具备推理功能。而这里介绍的查询算法是基于本体论的，在一般匹配失败的情况下，借助于本体论对常识性知识的刻划，系统能够寻找另一条合理的路径进一步查找，因此，它具有一定的智能性。假设一部分领域本体论如图3所示，我们来说明基于本体论的查询思想。

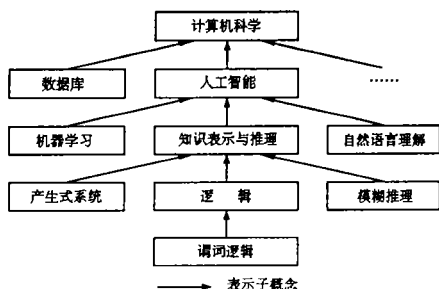


图3 一部分领域本体论

假设我们要查询懂得“知识表示与推理”的人，

但在数据库中没有相应的记录，按常理，懂得“知识表示与推理”子概念的人员也懂得本概念，因此，可以在“知识表示与推理”的子概念中继续查找。若仍找不到人选，我们找“知识表示与推理”的父概念，即“人工智能”，一般认为，懂得“人工智能”的人多少懂一些“知识表示与推理”。但在父概念上查询可能使查询结果的精确度下降，因此我们是在最坏情况下使用这一策略。以上查询思想可用算法1描述。

算法1：基于本体论的查询算法

输入：本体论，查询关键字；

输出：满足条件的记录；

方法：

```

begin
  result=φ;
  将查询关键字在本体论中找到对应的概念C;
  result=直接满足概念C的所有记录;
  if result 不空 then return result
  else begin
    result=满足概念C所有子概念的记录;
    if result=空 then
      result=满足C的父概念的记录;
    return result;
  end
end;

```

下面我们来形式化说明基于本体论的查询算法。

一个本体论可表示为一个带权有向图 $G=(V, E)$ ，其中， V 表示结点， E 表示有向边，其类型可以有多种，比如 $isSubFieldOF$ ， $hasCompetence$ 等等。有向边上的权是一个标注对象，它可以用以说明某人具有某种能力的强弱程度等。一个基于本体论的启发式表达式可表示为 n 个函数的组合运算：

$$f_1 \circ f_1 \circ \dots \circ f_n \quad n \geq 1 \tag{1}$$

其中， $f_i = (\lambda)^i \quad 1 \leq i \leq n$

λ 表示一条有向边 $e \in E$ 或其反向 e^{-1} ；

r 是一个偏序闭包，它表示路径长度，其取值形式可为： $1, 1 \dots m, \geq 1, *(表示 \geq 0)$ 或 $+(表示 \geq 1)$ ($1, m$ 为正整数或 $0, 1 \leq m$)。

公式1说明它以一有向图的结点集作为输入，对每个结点，沿着公式所说明的边以从右到左的顺序进行计算，每步产生的中间结果将作为下一步的输入。比如，公式

$$(hasCompetence^{-1})^1 \cdot (isSubFieldOF^{-1})^+$$

表示计算某概念的所有子概念，并求具备这些子概念能力的所有记录。

在设计具体查询时，我们可以给出不同的启发式公式——它将影响系统的查询结果。一个有启发式公式的查询算法由算法2所示。

算法2：基于本体论的启发式查询算法

输入：查询关键字集 ks ；启发式公式 $f_1 \cdot f_1 \dots f_n$ ；

输出：满足启发式条件的所有记录；

Windows环境下通用的图象透明处理算法

汪诗林 吴泉源

(国防科学技术大学计算机系 长沙 410073)

摘要 本文给出了一种用于在Windows环境下进行图象透明处理的通用算法, 该算法根据各种不同的Windows显示模式(如16色、256色、16位增强色和24位真彩色), 可以对任意颜色数(如各种单色、16色、256色和24位真彩色的图象)、任意存储格式(如Gif、BMP、TIFF、PCX等)的图象进行透明显示。

关键词 Windows 图象 透明处理

1 引言

在基于GUI(图形用户接口)的计算机软件系统中, 图象通常都是一种使用频率很高的媒体, 尤其是在多媒体应用系统之中。由于图象都是按照等宽、等高的格式存储的, 因此, 如果不作任何特殊处理, 我们看到的图象都是方方正正的“方块”图, 特别是在一个图象上叠加显示另外一个图象的时候, “方块”感尤为突出。比如我们要在一幅风景图上叠加显示一个小动物的图象(比如小鸟), 如果希望小鸟完全融入风景之中, 而看不出是叠加上去的, 就必须对小鸟图进行透明处理, 即只显示小鸟本身而屏蔽小动物之外的其它部分, 否则我们看到的将是一个关在“笼子”里的小鸟。

图象透明处理的实现原理其实并不复杂, 有很多

文章都已经作过介绍。尽管原理简单, 但在具体实现时, 却存在很多困难, 这是因为Windows可以工作在不同的显示模式下(如16色、256色、16位增强色和24位真彩色), 而图象也可以用各种不同的格式(如Gif、BMP、TIFF、PCX等)和各种不同的颜色数(如单色、16色、256色和24位真彩色)进行存储。常见的图象透明处理算法往往是针对某种具体的图象格式(如仅针对BMP图象), 不具有通用性, 对于其它不同的图象格式, 必须另外编写相应的图象透明处理算法。事实上, 不管图象的存储格式是什么、颜色数是多少(单色除外), 当它们被Windows调入内存以后, 它们在内存中的存储格式是完全相同的(以下称调入内存的图象为内存位图)。因此, 我们完全可以根据图象在内存中的存储格式来对该图象进行透明处理, 这样的透明处理算法将适应于任意存储格式和任意颜色数的图象。

由于在各种不同的Windows显示模式下, 内存位

收稿日期: 1999年6月7日

方法:

begin

(1)初始化堆栈stack为空;

(2)for i=1 to n do
push(stack, fi); //fi进栈

(3)temp=ks;

(4)while stack 非空 do //从右向左计算fi;

begin

(5) Fun=pop(stack);

(6) Temp=Fun(temp);

end;

(7) return temp;

end;

定理1. 算法2的时间复杂度 $O(en)$, 其中n为启发式公式中的函数个数, e为有向图边的条数。

证明: 算法第(6)步最费时间, 由于 $fi=(\lambda)^r$, 当r为*时(即求所有子树), 它要求遍历一棵子树, 若采用先深遍历算法, 其时间复杂度为 $O(e)$ 。算法第4步循环n次, 故整个算法时间复杂度为 $O(en)$ 。

5 结束语

关于本体论的研究还处于初期阶段, 基于本体论

的应用系统更是为数不多, 本文介绍的能力查询系统是作为这方面的一种尝试。该系统目前成功地运行于德国人工智能研究中心的Intranet上。进一步的研究工作是如何将该系统集成到整个企业知识管理系统中。

参考文献

- 1 D.O'Leary. Enterprise Knowledge Management. IEEE Computer, 31(3): 54-61, March 1998 PP54-61
- 2 T.R.Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. in Int. Journal of Human and Computer Studies, Vol.43 of 5/6, 1995 PP907-928
- 3 N.Guarino. Formal Ontology and Information Systems. In. N.Guarino(ed.) Formal Ontology in Information Systems, Proc. of FOIS'98, Trento, Italy, 6-8 June 1998
- 4 M.Liao, K.Hinkelmann, A.Abecker, M.Sintek. A Competence Knowledge Base System as Part of the Organizational Memory, in F.Puppe (ed). XPS99, 5th Biannual German Conf. on Knowledge Based Systems. Springer-Verlag, LNAI 1570, Wuezburg, Germany, 1999, PP125-137
- 5 JDBC 主页 <http://java.sun.com/products/jdbc/>. 1998
- 6 A.Abecker et al.. Toward a Technology for Organizational Memories. IEEE Intelligent Systems and Their Applications. 13(3), May/June, 1998