

严亚兰

## 论知识仓库的架构<sup>\*</sup>

**摘要** 在知识管理时代,知识仓库具有隐性知识共享和转化功能、知识存储和检索功能、知识分析功能、新知识产生和反馈功能,以及用户行为分析跟踪功能。知识仓库的架构分为6个主要构件:共享和获取隐性知识模块;获取显性知识模块;知识的抽取、转换和储存模块;知识分析模块;用户(系统管理员)界面模块;3个反馈环。目前,已经开发出一些知识仓库软件系统。图1。参考文献8。

**关键词** 知识管理 知识仓库 功能 架构

**分类号** G250

**ABSTRACT** In the age of knowledge management, knowledge warehouse has the functions of intangible knowledge sharing and transformation, knowledge storage and retrieval, knowledge analysis, new knowledge production and feedback, and tracking of user behaviors. In this paper, the author discusses major components of knowledge warehouse and some software systems for the construction of knowledge warehouses. 1 fig. 8 refs.

**KEY WORDS** Knowledge management. Knowledge warehouse. Function. Framework.

**CLASS NUMBER** G250

### 1 知识管理时代知识仓库的地位

知识可划分为显性知识(explicit knowledge)和隐性知识(tacit knowledge)两类。其中,显性知识是指能够用语言、符号、规则、公式或对象等正式表达并能够传输给他人的知识;隐性知识是深深根植于人脑中的信念、观点、创意和智力模型,包括某人长期从事某项活动或职业而形成的主观经验、洞察力和直觉。二者紧密关联,并与经济活动融为一体,成为当代社会发展的主要推进力量。

知识管理是通过共享和抓住隐性知识并将其转变为显性知识,筛选、存储、加工、检索、传递和利用显性知识,创新新的知识来增加社会价值的。这种实践活动可用知识螺旋(knowledge spiral)来描述。在每一个螺旋中存在4个阶段:共享隐性知识阶段、隐性知识转变为显性知识阶段、显性知识转变为新知识阶段和通过学习产生新的隐性知识阶段。每一次新的显性知识和隐性知识的产生便是知识螺旋的一次上升。相对于知识管理,信息管理注重显性知识或称编码型知识(codified knowledge)的搜集、存储、加工、检索、分析和预测,这方面的研究成果主要表现为数据仓库(Data Warehouse,简称DW)的开发和利用。数据仓库使企业能抽取、筛选、存储大

量的数据,对用户的检索进行有效而准确的反应,并为决策活动提供了强大的基础。然而,数据仓库中仅仅存储了决策者所需知识的一部分,企业绝大部分智力财富以隐性知识的方式存在于员工的大脑中。因此,数据仓库不足以满足对知识检索的需求。为了满足知识管理和知识决策的需求,可以对现存的企业数据仓库进一步扩充,成为满足知识管理需求的知识仓库(Knowledge Warehouse,简称KW)。知识仓库能够对不同类型的知识(显性知识和隐性知识)和不同形式的知识(纯文本、二进制对象、模型等)进行捕捉、存储、编码、组织和分析。另外,这些知识还包括元知识(关于知识的知识)和分析后产生的新知识。

### 2 知识仓库的主要功能分析

知识仓库强调为决策者提供一个提升知识管理流程全过程的智能分析平台,它在很大程度上依赖于信息技术的发展并被日新月异的信息技术所推动。信息技术包括人工智能、神经网络、专家系统、多媒体技术、数据库、信息可视化、机器学习、信息分析与挖掘等,其中OLAP(Online Analytical Processing,在线分析处理)与DM(Data Mining,数据挖掘)是基于数据仓库技术而发展起来的信息分析与挖掘

<sup>\*</sup>本文系教育部人文社会科学重点研究基地重大项目“信息可视化与知识检索”研究成果之一。

工具。OLAP 是验证型的,建立在多维视图的基础上,重在根据已有的模式将直接源自数据库中的不同信息源的大量相关信息联系起来,以给分析人员一个清晰、一致的视图。OLAP 主要有两个特点,一是在线性(online),即对用户的请求做出快速响应;二是可以应用多种统计分析工具、算法对数据进行多维分析(multi-analysis)。DM 是挖掘型的,建立在各种信息源的基础上,重在发现隐藏在大量原始数据深层中的对人们有用的模式(patterns)。被抽取的模式即知识,具备可信、新颖、有效、易于理解的特点。DM 有两个主要过程,即建立模型和预测未来结果。在这些过程中,可应用统计分析、神经网络、决策树、遗传算法、模糊逻辑、基于记忆的推理(memory-based reasoning)等技术。OLAM (Online Analytical Mining, 在线分析挖掘)是 OLAP 和 DM 相结合的产物。OLAM 兼有 OLAP 多维分析的在线性、灵活性和 DM 对数据处理的深入性,目前 OLAM 正处于研究之中,针对 OLAM 的发展驱动力和基本结构,以下几点是必要的:OLAM 建立在多维数据库和 OLAP 的基础之上,能对任何它想要的数据进行挖掘;用户对挖掘算法具有动态选择的权力;具有强大的基于超立方的挖掘算法;能够协调好执行效率与挖掘的准确性之间的关系;具有灵活的可视化工具和良好的扩展性。

信息技术的发展使知识仓库的功能实现变为可能,知识仓库应具备以下主要功能:

(1) 隐性知识共享和转化功能。即知识仓库应具备共享隐性知识和将隐性知识转变为显性知识的能力。这依赖于机器学习、神经网络、信息可视化、多媒体技术等。

(2) 知识存储和检索功能。即知识仓库必须提供数据仓库所拥有的全部功能,并具有更加丰富的知识表现形式。知识仓库应能有效生产、存储、检索、管理各种形式的知识。

(3) 知识分析功能。知识分析是一个非常复杂的过程,分析任务常常利用各类归纳和演绎的人工智能技术,如神经网络、数据处理的分组方法、统计、基因算法、基于案例的推理等。每一个任务在输入数据、执行参数和输出格式方面都有自己的要求。

(4) 新知识的产生和反馈功能。知识仓库中的知识随着不同的反馈环而得到实时更新,如通过头脑风暴法产生、共享和获取新的隐性知识;从用户刚刚验证和证实的结果中产生新的显性知识。

(5) 用户行为分析跟踪功能。知识仓库能够根据用户所提供的信息、用户的行为习惯和倾向进行跟踪并针对性地提供决策服务。

### 3 知识仓库的架构

知识仓库包括 6 个主要构件(见图 1):共享和获取隐性知识模块;获取显性知识模块;知识的抽取、转变和存储模块;知识分析模块;用户(系统管理员)界面模块;3 个反馈环。

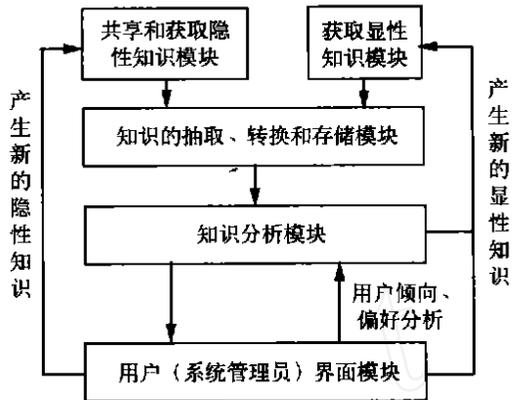


图 1 知识仓库的架构

#### 3.1 共享和获取隐性知识模块

隐性知识是无法交流的知识。例如,学生通过书本可间接地学到用文字表达出来的显性知识,而通过上课不仅可学到显性知识,而且能体会到教师讲授中无法表达的隐性知识。这即是所谓的言传身教,但在对所讲所学知识的理解上学生和教师总存在差别。又如,邀请一位有经验的管理者作报告,假设该管理者对自己所理解的管理理论和实践经验尽其全力进行表达,是否听报告的人在报告后便能成为与其一样出色的管理者?答案是否定的。究其原因,该管理者能表达出来的知识是显性知识,而他头脑中大量存在的是他无法表达清楚的隐性知识,这些隐性知识在该管理者的管理实践中起非常重要的决策支持作用。知识管理非常强调对隐性知识的挖掘。共享和获取隐性知识模块应具体拥有以下功能:行为隐性知识获取。表现为行为跟踪、动作慢放,如将外科手术、工艺操作等过程进行录制生成多媒体信息,再对这些多媒体信息进行慢放,通过观摩这些慢放的画面,可以共享人的行为中所拥有的隐性知识。提供一个平台,让大家各抒己见,如在 BBS、聊天室或头脑风暴法基础上加上跟踪软件,分析表达出来的、看起来分散而孤立的显性知识之间的关系,挖掘隐性知识。基于模型环境的数学模型抽取。基于专家环境的规则抽取。

#### 3.2 获取显性知识模块

这个模块功能类似于数据库中相应的功能,能够对显性知识进行搜集和筛选。

### 3.3 知识的抽取、转换和存储模块

这个模块是知识仓库的基本构件。它是一个面向对象的知识库管理系统(Knowledge Base Management System,简称KBMS),集成了知识库、模型库和分析任务等。知识以框架、规则、语义网络等形式存储在知识库中。KBMS管理如何将广泛的知识对象集成为一个功能体,这些知识对象包括数字、文本流、验证的模型、元模型、电影剪辑、动画片断以及控制它们的软件。KBMS在面向对象的环境中实施,不仅管理数据,而且管理所有的对象、对象模型、过程模型、案例模型、对象交互模型和用来处理、解释知识并产生知识库的动态模型。

### 3.4 知识分析模块

知识分析模块处理所有与分析任务有关的活动,包括知识工程、任务控制、判断生成和技术管理。知识工程次模块是基于专家系统的子系统,它负责开发知识分析目标和基本的模型环境原则。

任务控制次模块处理所有数据要求和运行中的交互(包括对用户倾向偏好的分析和跟踪),即以任务控制台的形式作为任务交互的基于人工智能的代理决策者,使真正的决策者不必知道其中的技术、微小区别和交互过程等。

判断生成次模块对各类分析任务的输出(包括基于用户行为跟踪分析而提供的针对性决策服务)进行评估,特别是对于因果任务,筛选出基于某个标准的假的或不一致的结果,然后产生希望能使用户明白模型环境的简单而深刻的解释。在这个过程中,判断产生模型与知识库、例子库和模型库交互,利用演绎知识、类比推理和其他技术。

技术管理次模块对分析技术库进行管理,它提供从新生分析算法到对象模型类的封装、遗留数据挖掘集成应用、新生分析模型和元模型的对象模型库合并,等等。

### 3.5 用户(系统管理员)界面模块

这个模块处理KBMS和用户间的所有分析通信,包括5个功能子模块:判断界面、输入处理器、输出处理器、在线帮助和系统管理员界面。

判断界面是为了有效帮助用户确定一个或多个条件进行观察,它包括一个通过变化一个或多个参数值来提出相关例子的类比构件,还包括一个或多个交互图形以显示操作过程,以便用户能快速看清什么已经试过、哪些例子可引向更深的洞察。判断界面还应能提供基于计划分析任务的具有潜在价值的案例分析。

输入处理器在用户和分析任务间提供界面,它将用户确定的自然语言、SQL检索语句等转变成机器可执行的检索,或将用户的语音和绘图等信息转变成机器可识别的数据。

输出处理器为每一个分析结果选择最合适的可视化显示,如图、自然语言、多项式、决策树等。这种选择是基于分析任务输出和用户偏好的结合,而这又是基于适应性机器学习算法,它可分析当前用户的历史行为。

帮助子模块向用户提供有关模型(如参数范围、测量单位、内部模型结构)、例子(与基本案例的差异、关键决策变量值)、相关知识(如元模型、元数据、基本原则、分析目的)和分析任务(如应用技术、技术描述、结果的解释跟踪、使用的技术参数、技术的优势和局限)的信息。

系统管理员界面是为系统维护者提供的一个与知识库通信的端口。

### 3.6 反馈环

知识仓库包括3个反馈环,一是从用户(系统管理员)界面模块到知识分析模块之间的环,它通过分析用户的倾向偏好来更好地为用户提供知识分析及其结果输出服务;二是从知识分析模块和用户(系统管理员)界面模块到获取显性知识模块之间的环,知识分析和用户产生新的显性知识,新的知识通过这个环存储到知识库中,这是知识管理的第3个阶段;三是从用户(系统管理员)界面模块到共享和获取隐性知识模块之间的环,用户通过学习而产生新的隐性知识,这是知识管理的第4个阶段。

## 4 知识仓库系统开发实践

目前已出现支持知识仓库部分功能的软件系统。下面针对知识仓库应具备的功能对其中的几个软件进行分析。

首先,知识仓库应当有效地生产、组织、检索和管理各种形式的知识,具有智能模型抽取、知识工程等功能,向决策者提供一个智能分析平台,以提升知识管理全过程。下面以IBM software Solution为例进行分析。IBM是世界最大的信息技术企业,自20世纪80年代以来一直处于帮助企业进行革新的领先地位,其软件提供各类计算机平台的最广范围的应用、中间件(middleware)和运行系统。IBM的KnowledgeX技术能帮助企业通过更好地获取、分析和共享数据,做出更精明的决策,能平衡所有类型的数据,而不论这些数据是来自数据库内,还是来自公开发表的新闻或者因特网。IBM计划未来将KnowledgeX技术集成到DB2广义数据库(DB2universal Database)

和企业智能方案 (business intelligence solutions) 中,以为客户带来更多的功能。通过镜像人类解决问题的过程, KnowledgeX 从各类数据源中搜集信息,发现数据间的关联和关系,将数据应用到问题中,并帮助开发出方案。目前利用 KnowledgeX 的一些领域包括竞争情报、合并和获取、销售力自动化、诉讼、垂直产业应用和知识管理。当今的企业正在利用更多的数据源以构建或保持其竞争优势——来自 Internet 和 Extranets,或来自文档和公司数据库。将 IBM 的数据库和企业智能方案与 KnowledgeX 结合起来,企业将能更好地处理各种来源的数据,寻找获胜机会,并在企业中将知识转变为行动。KnowledgeX 软件能揭示企业数据中隐藏的关系,决定这些关系的力量,并识别它们对企业目标和问题的实用性。KnowledgeX 能显示关键产业活动者间的关系,解释过去的行为、合同或联盟,能分析并链接看起来分散的信息以跟踪发展方向并智能地预测竞争者未来的行动。KnowledgeX 技术对 IBM 在数据和企业智能空间如 DB2 广义数据库,文本智能挖掘机 (intelligent miner for text), 可视化仓库 (visual warehouse) 等中的工具和方法起到补充作用。KnowledgeX 与企业智能工具如数据仓库、数据挖掘和在线分析处理一起允许企业从数据源中获取更多的洞察力。目前,全球的客户正利用 IBM 企业智能系统来更快、更明智地决定应进入哪个市场、雇用哪个客户、宣传哪个产品,以达到提升利润和竞争力的目的。

其次,知识仓库应当能执行和管理各类知识分析任务及其支持技术(如敏捷度分析、机器学习、神经网络、知识仓库挖掘和模式识别等)。下面以 VantagePoint 为例进行分析。VantagePoint 提供从文本数据库中挖掘知识的能力。它具有以下功能:抽取并索引新数据。通过检索,用户将数据引入到 VantagePoint 中,在这个过程中,VantagePoint 使用了一个数据库构成文件(这对于不同的数据库和数据库提供商都不一样),利用模式匹配、规则基础和自然语言处理技术从文本中挖掘字段,除记录通常的字段以外(如作者、单位和分类号),VantagePoint 还能利用自然语言处理技术从文摘中抽取有意义的词和词组。同发生矩阵。利用同发生矩阵,用户可对表进行交叉。如利用同发生矩阵中的“作者”和

“出版年代”,用户可观察某作者一段时间内的出版倾向,并按年代浏览他的作品,这可提供有关哪个作者在某领域工作时间最长,现在他们是否还活着等方面的信息。表或同发生矩阵中的数据可输出到其他应用中以便进行可视化显示或分析。映射/网络。在一维(表)和二维(同发生矩阵)分析之外,VantagePoint 还可运行多维统计分析以识别概念之间的簇和关系。Vantagepoint 可将数据分解为更小的、分散的数据集以进行识别。VantagePoint 利用模糊匹配技术进行数据识别、关联并减少数据冗余。用户可创建、编辑同义词词典。用户可利用 VantagePoint 的微软 VBScript 扩展功能来调整操作。VantagePoint 在 VBScript 中增加了 5 个 VB 对象和 50 多种 VB 方法,允许用户创建和分配脚本,这可使定制分析技术重复使用。

最后,知识仓库应具有用户友好的界面和共享隐性知识的平台,下面以 Lotus Developments 为例进行分析。Domino.Doc 将小范围专家使用的文档管理界面扩展为基于网络环境的、更灵活的、适用于企业内每一个成员的解决方案。企业成员通过 Web 浏览器共享信息和应用,企业服务器通过扩充能力、灵活性和安全性的平衡使其在知识管理中发挥主要作用。

#### 参考文献

- 1 Hamid R. Nemati et al. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems* 33(2002), pp. 143 ~ 161
- 2 马海群. 论知识经济、知识管理与知识产权. *图书情报知识*, 2000(2)
- 3 褚峻. 知识管理的资源性分析. *图书情报知识*, 2000(2)
- 4 刘夫涛. 从 OLAP、数据挖掘到 OLAM(上). *网络世界*, 1999 - 02 - 01
- 5 刘夫涛. 从 OLAP、数据挖掘到 OLAM(下). *网络世界*, 1999 - 02 - 08
- 6 <http://www.software.ibm>
- 7 <http://www.thevantagepoint.com>
- 8 <http://www.lotus.com>

严亚兰 武汉大学信息管理学院博士研究生。通讯地址:武汉市。邮编 430072。(来稿时间:2002-08-29)