

知识管理与知识共享

胡舜耕 王克宏

(清华大学计算机科学与技术系 北京100084)

Knowledge Management and Knowledge Sharing

HU Shun-Geng WANG Ke-Hong

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract Knowledge management aims at knowledge innovation and knowledge sharing, and focuses on added value and added services of knowledge. In this paper, concepts, key technologies and applications of knowledge management are introduced, and relevant problems are discussed.

Keywords Knowledge management, Knowledge sharing, Knowledge storage, Knowledge retrieval, Knowledge presentation, Ontology, Application

现在的人类社会,知识无处不在,无处不用。越来越多的企业意识到人与知识是企业竞争力的根本,于是,“以人为本”成了许多企业文化的核心,“知识资本”和“知本家”体现出人们对知识价值的认同。根据报道,财富1000中近半数的企业设立了知识主管,负责为知识共享创建基础设施和文化环境,为实施知识管理创造条件。不仅工业界,学术界也对知识管理倍加关注。本文介绍知识管理与知识共享的概念,关键技术及其应用,对相关问题进行分析和探讨。

1 知识与知识管理的概念

自从有了智能的人类就有了知识,而语言文字给知识的广泛传播和共享创造了条件。虽然知识是个很古老的话题,但人类对究竟什么是知识并未达成共识。

我们认为,知识是人们在反复实践中形成的对客观事物及其运动过程和规律的认识。在反复实践和认识的过程中,人脑形成对事物本质的认识,构成头脑中的主观知识,主观知识也称隐性知识。隐性知识经过各种载体表达出来,并通过人们的实践检验则成为客观知识,客观知识也称显性知识。人们通常所说的知识是指显性知识,然而,获取隐性知识是创造显性知识的必由之路。

二十世纪八十年代以来,出现了以信息技术为主、包括了生命科学、基因工程、新能源、新材料、空间技术、海洋开发技术、环境科技等的“高新技术”。由于高新技术的产业化,经济社会日渐出现一些新的特征,与传统经济相比,发生了质的变化。1990年,联合国研究机构提出了知识经济的说法,明确了这种新型经济的性质。1996年,经济合作与发展组织明确定义了“以知识为基础的经济”(knowledge based economy),首次提出了这种新经济的指标体系和测度。

伴随知识经济的萌芽,知识管理备受学术界和工业界的关注^[1-4]。但是,就象人们对知识有不同的理解一样,以知识和人为管理对象的知识管理,目前也没有一个广泛认同的定义。兹列举知识管理的几个定义如下。

按照美国戴尔集团创始人之一卡尔·弗拉保罗的说法,

“知识管理就是运用集体的智慧提高应变和创新能力”,是为企业实现显性知识和隐性知识共享提供的新途径。

美国“知识的进化”作者 Verna Allee 对知识管理的定义是:“帮助人们对拥有的知识进行反思,帮助发展支持人们进行知识交流的技术和企业内部结构,并帮助人们获得知识来源,促进他们之间进行知识的交流。”

Daniel E. O'Leary 认为:“知识管理是将组织可得到的各种来源的信息转化为知识,并将知识与人们联系起来的过程。知识管理是对知识进行正式的管理,以便于知识的产生,获取和重新利用”。这种解释着重阐明了信息、知识和人在知识管理过程中的不同角色。

Wiig 指出,知识管理主要涉及四个方面:自上而下监测和推动与知识有关的活动,创造和维护知识基础设施,更新组织和转换知识资产,使用知识以提高其价值。

P. Quitas 等将知识管理看作“是一个管理各种知识的连续过程,以满足现在和将来出现的各种需要,确定和探索现有和获得的知识资产,开发新的机会。”知识管理的目标包括六个方面:第一,知识的发布,以使一个组织内的所有成员都能应用知识;第二,确保知识在需要时是可得的;第三,推进新知识的有效开发;第四,支持从外部获取知识;第五,确保知识、新知识在组织内的扩散;第六,确保组织内部的人知道所需的知识在何处。

尽管上述定义对知识管理的理解不尽相同,但都隐含了知识管理包括的三个基本要素:知识生产过程管理,知识传播过程管理和知识使用过程管理。因此,我们认为知识管理是对知识的生产、传播和使用的全程监控。它以信息技术为基础,为知识的生产、传播和使用提供开放的可管理的协同工作机制。

2 知识管理与信息管理的关系

知识与信息,知识管理与信息管理是紧密相关的。信息是“事物运动的状态及其变化的方式^[5]。”信息有时与主体有关,而知识是客观的,与主体无关。知识是信息,是信息的提炼和

胡舜耕 博士后,副教授,感兴趣的研究方向为 Agent 计算,智能信息处理,知识管理,电子商务,电子政务。王克宏 教授,博士生导师,主要研究方向为网络计算与知识工程,基于 Web 技术的知识挖掘,基于 Web 技术的软件技术研究与开发,嵌入式技术。

一般化,然而,信息不一定是知识。

信息管理是为实现组织目标、满足组织需求而对信息资源进行的规划、开发、集成和利用。在知识管理与信息管理的不关系方面,信息管理是知识管理的基础,知识管理是信息管理的拓展和延伸。信息管理和知识管理之间的区别在于:

首先,知识管理拓展了信息管理的对象。知识管理的对象包括知识以及传统上知识的创造者一人,而信息管理仅仅局限于用语言、文字、图形、图像、影视等各种载体来表示的信息。因为知识存在于人的头脑之中,所以只有人能够充当知识创造的主体,尽管计算机有强大的信息处理能力,它只能是知识创造的辅助工具。因此,知识管理把人纳入了其管理的范畴。

其次,知识管理提升了信息管理的地位和作用。信息管理服务于组织的生产、经营和管理。信息管理系统是为组织整体的管理、控制和决策等服务,它的成功运行可以保证组织运作的高效率、及时性和适应性。在知识型组织中,知识的生产、传播和利用是组织的主要工作之一,是组织核心竞争能力的综合体现,是组织在激烈的全球竞争环境中制胜的法宝。知识管理不仅仅是人与人之间的知识共享,还包括知识的创造和广泛利用。创造知识的增值价值和增值服务是知识管理的目的,也是它吸引工业界的原因所在。

再次,知识管理强化了对信息管理基础设施的要求。特别是在知识传播方面,要求做到在任何时间,任何地点,任何人都可以获得所需要的知识,以保证知识服务的实时性和高效性。

最后,知识管理是与信息技术,人工智能,管理科学,人文科学和经济学等相关的一个边缘学科。知识管理研究人和知识在知识型组织中的地位和作用,研究人类生产知识的机制,以及信息技术如何被用来提供知识生产、流通和使用的支撑环境。知识管理还研究知识如何有效地组织、存储和处理,以协助乃至代替人类使用和创造知识。与之相比,目前的信息管理在相当程度上还属于计算机应用技术的范畴。

3 知识管理关键技术

知识管理要实现两个基本目标:第一,已有知识的共享与重用;第二,创造新知识。它是跨越某一时段的实践过程,与人、商业应用和信息技术密不可分。

知识管理过程包括若干阶段:

(1)知识聚集:包括已有知识的集成和新知识的获取。知识聚集又称为知识生产。

(2)知识的组织和存储:给获得的知识赋予一种结构并合理地存储,以便有效管理和使用。

(3)知识演变:由于知识聚集过程中发生的偏差,以及知识的时变性,需要更正、更新、删除旧知识,增加新知识。

(4)知识传播:使需要知识的任何个人和组织可以在任何时间、任何地点获得知识。

(5)知识使用:在知识处理系统之间和相关人员之间实现知识共享与重用。

知识管理离不开一个知识管理系统的支撑。知识管理系统不是一个单纯的知识发布系统,而是一个交互式的开放的协同工作环境。知识管理涉及许多技术,简述如下。

3.1 信息技术

与计算机互联网、电话网和电视网等相关的信息技术主要用于解决知识的快速、高效、实时和准确的传输,是知识传

播的基础设施。知识传播是集语言、文字、图像和声音于一体的多媒体数据传递,信息技术的快速发展为这种多媒体数据的快速、实时和准确地传递创造了条件。

3.2 Web 技术

目前,Web 技术及其多层体系结构是知识管理系统普遍采用的实现技术。Web 全称是 World Wide Web,缩写为 WWW,是一个基于 Internet/ Intranet 的、全球连接的、分布的、动态的、交互式超媒体信息系统,通过它可以访问遍布于全世界 Internet 主机上的链接文档。Web 技术主要涉及到两类标准,一个是应用层协议,称为超文本传送协议(HTTP);另一个是表示层句法,包括超文本标记语言(HTML)和扩展标记语言(XML)。此外,动态页面技术可以使得用户能够访问存储在数据库中的实时数据。

3.3 知识存储技术

W. H. Inmon 在“Building the Data Warehouse”一书中定义:数据仓库是面向主题的、整合的、稳定的、并且时变的数据集合,用以支持管理决策。它从多个内容相关、物理和逻辑上独立的数据源中提取面向主题的数据信息。数据仓库与事务数据库的区别在于:

(1)数据是面向主题组织的,不依赖于具体应用。

(2)集成多个异构数据源中的数据。

(3)数据仓库支持决策分析等联机数据处理,事务数据库主要是为了高效地获取事务数据。

(4)数据仓库不是一个事务一个事务地更新,而是整个数据库周期性地更新。

(5)数据仓库反映的是历史信息,而不是处理联机数据。每一数据项对应于某一特定时间。数据仓库的数据时限一般在5至10年,甚至更长。

(6)数据仓库容量大,可以存储海量数据,并且支持并行和分布处理。

知识管理需要管理的知识量是非常大的,这些知识需要有效地组织和存储。为了使知识能够更好地共享和重用,知识和知识处理系统应该分离,知识应该独立于知识处理系统。而且,知识通常是比较稳定的,更新的速度较慢,因此,知识库的知识时限达到几年甚至几十年。并且,知识库面对的主要问题不是查询,而是知识的分析和处理。所以,与事务数据库相比,数据仓库更适合于用来存储知识。

另外,现在越来越多的知识用 HTML 和 XML 文件的形式存储在 Web 服务器中,这是 Internet 给知识存储提供的新途径。

3.4 知识获取技术

Feigenbaum 说过:“知识获取是人工智能研究里最重要的中心问题,是人工智能研究中的关键。”知识获取就是从纷繁的信息当中发现、提取和挖掘知识。知识获取可以分为人工获取,机器辅助的人工获取,机器自动获取三种类型。就目前的技术发展而言,机器辅助的人工获取既能保证一定的精度,又能确保较高的效率,而知识的机器自动获取是人工智能追求的目标。

机器学习是机器获取知识的主要方法,而数据挖掘和 Web 挖掘是机器学习新的重要研究方向。

数据挖掘又称为数据库中的知识发现(KDD: Knowledge Discovery in Database),是从大量原始数据中挖掘出有用的、潜在的信息和知识(如概念、规律、规则、限制、模式、约束)。数据挖掘过程通常包括数据准备、预处理、数据发现,结

果表示和解释等阶段。数据挖掘方法众多,如利用 ID3 和 C4.5 等算法的信息论方法;粗集方法和覆盖方法等集合论方法;神经网络和遗传算法等仿生学方法;统计分析方法等等。数据挖掘的成功应用是在线分析处理(OLAP: On-Line Analytic Processing),它利用原本为统计而搜集的数据发现各种模式,分析变化趋势,进行预测,以及支持决策。尽管数据挖掘取得了可喜的成果,但学术界与工业界的沟通还不够,学术界的研究成果应该更多地用于解决工业界的实际问题。

Web 内容挖掘、Web 结构挖掘和 Web 使用挖掘是 Web 挖掘的三个组成部分。Web 内容挖掘通常指 Web 文本挖掘。TFIDF 和 Naive Bayes 方法能对大量文本进行快速、有效的自动分类。G-HAC 和 k-means 是文本聚类的代表性算法。Web 结构挖掘和 Web 使用挖掘方面也有一定研究基础^[6]。

目前 Web 挖掘存在如下不足:第一,Web 内容挖掘、Web 结构挖掘和 Web 使用挖掘相互独立,没有互相利用,有机结合。Web 内容挖掘算法没有充分利用 Web 文档中的结构信息和超链信息。第二,Web 内容挖掘几乎没有利用语义信息。目前的 Web 内容挖掘大都是基于信息提取的,而没有在获取 Web 文档的语义信息基础上建立挖掘的理论和算法,这使得挖掘结果的精度较差,质量不理想。第三,Web 挖掘算法尚不能有效处理海量数据。这也是传统数据挖掘面临的难题。

3.5 知识表示与 Ontology

知识是知识管理的管理对象之一,是人类认识客观世界的创造性成果。在人类思维活动中,人是加工和处理知识的主体,知识以某种形式存在于人脑之中,或以文字等形式记载下来,供人类学习使用。因此,知识构成人类思维活动的环境,是人类创造性工作的基石。

机器要加工和处理知识首先必须拥有知识,因此,一个关键性问题是机器如何表示知识。就好像人类创造了语言文字,使知识得以保存记录下来,这些知识既是人类创造的成果,又是后人学习和再创造的基础,知识表示就是机器描述知识的语言和文字。

1991 的 ARPA 知识共享努力设想智能系统建造的一个新方法:聚集可重用构件。今天,建造基于知识的系统通常必须从头开始建立知识库。其实,它可以通过聚集可重用构件来完成。系统开发人员只需创建与他们的系统任务相关的特定知识和推理机。新的系统将和已有系统进行互操作,利用它们执行某些推理。通过这种方式,知识、问题求解技术和推理服务可以在系统之间共享。这种方法有利于廉价地建立更大和更好的系统。然而,时至今日,ARPA 的设想还远未实现。其中,一个重要原因就是知识不能共享与重用。

Angus 等指出:目前的知识管理不能重用从经历中获得的知识,因为它不以一种形式化的方式共享。知识的共享与重用是知识管理要求实现的目标之一。可见,知识共享与重用不仅是智能系统建造中的普遍性问题,更是知识管理面临的紧迫研究课题。我们认为,知识共享与重用需要一种形式化的共享的知识表示规范。尤其是现在的知识管理系统乃至智能系统都是分布系统,集中式知识表示已经不能满足应用的需要,必须研究知识的分布表示方法。知识表示的研究应该从基础开始,这个基础就是 Ontology,它是共享知识表示的关键。

那么,什么是 Ontology 呢?学术界对 Ontology 并没有达成共识,引用较多的是 T. R. Gruber 的定义。他认为,Ontology 是一个共享的概念化规范,而概念化是指某个领域中的概念及其相互关系,是我们希望描述的世界的一个抽象的简化

的视图。

为什么说 Ontology 是共享知识表示的关键呢?原因是:第一,Ontology 阐述了知识的结构。给定一个领域,它的 Ontology 构成了面向这一领域的知识表示系统的核心。没有知识之下的 Ontology 或领域概念化,就没有一个表示知识的词汇。第二,Ontology 使知识共享成为可能。假设我们做了一个分析,得到某个知识领域的一个满意的概念化集和它们的代表项。为了建立知识表示语言,我们在项与 Ontology 中的概念和关系之间建立联系,设计用概念和关系编码知识的语法。因为有了 Ontology,就不必重复知识分析过程,而可以与他人共享这一知识表示语言。共享 Ontology 形成了领域相关的知识表示语言的基础。基于 ontology 构造的知识表示语言是内容丰富的(或语义丰富的),它们有大量包含复杂领域内容的项。共享 Ontology 使我们能够建立描述特定情形的特定知识库,这种类型的共享将大大增加知识重用的潜力^[7]。

Ontology 与通常所谓的词典是有区别的:

- (1)词典是人读的,Ontology 必须方便机读。
- (2)词典注重概念本身的描述,Ontology 既注重概念本身的描述,又注重概念之间关系的表示。
- (3)词典通常以自然语言描述,Ontology 可以用其它符号语言描述。

经过多年的努力,研究人员已经建造了一些有影响力的 Ontology。Wordnet 是一个基于语言心理学原理的英语词汇数据库,可用作在线词典,推理词典和分类词汇数据库。CYC 是世界上最大的最完善的常识知识库。Sensus 是一个主要用于机器翻译的 Ontology 实例。值得一提的是,董振东先生积多年的心血建造的 HowNet(“知网”)是我国一个不可多得的共享 Ontology。

建造 Ontology 需要一种表示语言作为基本的形式化工具。原则上,Ontology 能够用各种各样的语言实现:非形式化的,半形式化的或形式化的,并无强制性的规范。目前,有两种逻辑常被用作 Ontology 的形式化工具:一,谓词演算及其变种;二,框架逻辑。谓词演算及其变种较为常见。

M. R. Genesereth and R. E. Fikes 提出了 KIF (Knowledge Interchange Format),它是一种基于扩展的谓词演算的形式语言,用来表达领域事实知识。Frame Ontology 是一种框架表示语言,它包含类、子类、属性、值、关系和公理等原语。Ontolingua 中的 Ontology 就是使用 Frame Ontology,其语法是基于 LISP 的。CYCL 是 CYC 的知识表示语言,类似于一阶谓词演算。XOL (XML-Based Ontology Exchange Language) 是一种 Ontology 交换语言,提供交换 Ontology 的定义格式,有基于 XML 的语法。OIL (Ontology Inference Language) 是基于框架的表示语言,还在不断完善之中。

尽管 Ontology 常用来建立系统、人和组织之间的通信,协助系统之间的互操作,支持知识系统的设计和开发,但真正用 Ontology 来建立领域知识模型的应用却很少。建立可以共享与重用的领域 Ontology 不仅是知识表示的需要,也是实现知识共享与重用的关键,是实施知识管理的一项基础性工作。

3.6 Agent 技术

Agent 是有智能的实体,它具有自主性、社会性、反应性、自适应性等特征。多 Agent 系统是一个松散耦合的 Agent 网络,这些 Agent 通过协作解决超出自身的知识和能力的问题。所谓松散耦合意味着 Agent 把更多的时间花在业务处理而不是通信上。多 Agent 系统中的 Agent 是自主的,它们可

以是不同的个人,采用不同的设计方法和计算机语言开发而成的,因而可能是完全异质的。多 Agent 系统涉及多 Agent 系统的体系结构,Agent 通信语言,Agent 通信协议,协调与协商技术和多 Agent 学习等关键技术^[8,9]。

Agent 在 Internet 中的应用越来越深入,Agent 与 Web 不断融合。而且,专家预言,在本世纪,基于 Agent 的软件工程将取代面向对象的软件工程而成为软件工程的主流。我们相信,Agent 技术在知识管理的组织建模,体系结构,知识处理和知识搜索等方面有广泛的应用前景。

4 知识管理的应用

我们先看看 Microsoft 和 IBM 提供的知识管理产品和服务。

自从盖茨的《未来时速》中提出了“数字神经系统”概念之后,Microsoft 将数字化、知识管理和电子商务列为关注的三大主题。Microsoft 的知识管理系统平台为“Tahoe”。它是集文档管理、文档索引/检索和协同工作于一身的企业门户。Tahoe 的文档管理包括版本控制、文档的作者与密码属性管理、文档发布控制、签发控制等功能。在文件索引方面,Tahoe 可以进行全文检索,也可以对网站、文件系统、Exchange 服务器、Lotus 服务器等多种信息源进行检索。

IBM 以 Lotus 现有技术为基本出发点,制定出独特的理论框架,并确立了知识管理产品策略,K-Station 企业门户和其配套产品 Discovery Server 是 IBM 创新的知识管理解决方案的重要组成部分。在 K-Station 中,每个人都有自己的场所一个人场所(Personal Place),沟通场所(Community Place)为由相关人员组成的小组提供了共享与共同工作的环境。在场所中可以看到何人正在线上,并列出生共享场所的清单,在线上的人可以相互进行即时的消息沟通。Discovery Server 具有先进的数据搜索和专家定位功能,用于帮助企业用户高效率地维护和扩大知识资产。K-station 与 Discovery Server 相结合时,能够组成功能强大的 Lotus 知识发现系统,可以帮助企业用户高效率地发现、分类和管理企业知识,将知识转换为有价值的资产。

IBM 的知识管理研究院(IKM:The Institute of Knowledge Management)确定了九个研究领域,每一个领域内主要进行一些最基础性的研究,以求建立未来跨国公司的学习框架。它可能是大型企业里一个最系统化的知识管理研究计划。九个研究领域是:团体和团队,知识战略,专家网络管理,客户知识,技术的目标,知识经济,创新,灵活性和响应能力,社会资本。

知识管理首先是一种管理理念,通过知识管理平台将这种理念贯穿于对知识与人的管理过程之中。信息管理是知识管理的基础,知识管理是信息管理的拓展和延伸。知识管理不会取代信息管理,信息系统和知识系统将通过信息和知识的共享与重用形成共存的局面。

知识管理在企业信息化,商业智能(Business Intelligence)和电子政务(Electronic Government)中有广泛的应用。

企业信息化包括 CAD/CAM/CAPP(计算机辅助设计/计算机辅助制造/计算机辅助工艺流程)、MIS(信息管理系统)、MRP(制造资源计划)、ERP(企业资源计划)、CIMS(计算机集成制造系统)、EC(电子商务)等在企业中的广泛应用。企业信息化的目标是,在企业产品开发、生产、销售和服 务的全过程中,通过信息集成、过程优化、资源配置,实现物流、信息流、资金流的集成和优化,提高企业市场应变能力和竞争能力。商业智能针对现有与未来的数据做采集、挖掘、整理、分析,以期对已有的数据有更多的了解,以便提供风险管理,决策支持和商机策略,提高企业竞争力,增加市场份额。电子政务的根本目标是政务的自动化、电子化和网络化。在政务活动中,信息和知识是决策的依据。可见,企业信息化,商业智能和电子政务等都是信息和知识的应用,知识管理的引入能够更好地支持这些应用,为它们提供宽广实用的创新知识,甚至有可能产生更高级的知识应用形态服务于各种实际需求。

此外,传统的智能系统本身就是知识系统,知识管理给智能系统带来新的发展契机。知识管理将推进知识的共享与重用,并首次关注知识创新以及知识创新的主体一人,这为智能系统走向分布、开放和自主创造了前提条件。可以预期,与知识管理相结合,将出现新型的智能系统。

结束语 知识管理首先来源于企业发展竞争力的需要,因此,Microsoft 和 IBM 的知识管理产品和服务都是面向企业的。但是,凡是需要知识的场合就需要知识管理,知识管理的应用远比企业应用范围宽广得多。可以说,知识管理还刚刚起步,它的理论、技术和应用有待拓展和深化。

参 考 文 献

- 1 Available at: <http://www.kmresource.com/>
- 2 O'Leary D E. Enterprise knowledge management. IEEE computer, March 1998. 54~61
- 3 李啸尘. 新人力资源管理. 全四卷. 中国:石油工业出版社, 2000
- 4 Available at: <http://www.chinakm.com/>
- 5 钟义信. 信息科学原理. 中国:福建人民出版社, 1990
- 6 Kosala R, Blockeel H. Web mining research: a survey. In: Paul S. Bradley, Sunita Sarawagi, Usama M. Fayyad. SIGKDD Explorations. Boston, MA, USA: June 2000
- 7 Chandrasekaran C, Josephson J R, Benjamins V R. What are ontologies, and why do we need them? IEEE Intelligent Systems, Jan./Feb. 1999. 20~26
- 8 胡舜耕. 多 Agent 系统规范和若干关键技术及其在自动文摘中的应用研究:[北京邮电大学博士研究生学位论文]. 中国:北京邮电大学图书馆, 2000
- 9 胡舜耕,张莉,钟义信. 多 agent 系统的理论、技术及其应用. 计算机科学, 1999, 26(9): 20~24