

基于本体论的知识管理模型及其实现

章婕

(大连海事大学 计算机科学与技术 辽宁省大连市 中国 116026)

摘 要 为了将大量无序的信息转化为企业有用的知识,以提高企业的竞争力,文中提出了一种基于本体论的知识管理结构模型,接着从知识的俘获和存储及知识的检索两个方面,分析了模型的实现技术。由于/O1在该知识管理模型实现中的关键作用,文中重点探讨了这方面的内容。该模型解决了知识管理中,知识的俘获、存储与检索问题。

关键词 可扩展标记语言; 文件类型定义; 知识管理; 本体论; 组织记忆

中图法分类号 TP311.90 DOI号 10.3969/j.issn.1001-3695.2015.04.033

Model of Ontology-based Knowledge Management

Zhang Jie

(Computer science and technology, Dalian maritime university, Liaoning Dalian, 116026, China)

Abstract In order to improve enterprise's competitive ability, vast out-of-order information need be available knowledge. Therefore, a model of knowledge management framework is presented in this paper. After that, the implementation technique of this model is analyzed from two aspects: knowledge capturing, saving and knowledge saving. XML technique, which is applied in the model, is discussed especially. This model solve the question of knowledge capturing, saving and searching in knowledge management.

Key word XML; knowledge management; ontology; organizational memory;

1 引言

随着信息技术的普及和深入应用,企事业单位已将大量关于顾客、产品开发和协作的知识包含于以电子文档、电子表格、Web页面和E-mail等形式的信息体,存储于内部的信息系统;同时,因特网上也出现了与日俱增的令人感兴趣的信息(主要表示于Web页面),从中获取有用的知识,可增加技术、竞争、商务和策略智能。面临经济的全球化和激烈的市场竞争,如何有效地重用包含于企事业内外信息系统的相关知识,去提高运营效益和市场竞争能力,已引起越来越多的知识密集型企事业单位的关注和重视。

于是,知识管理(KM, Knowledge Management)技术应运而生。Y0就是指综合利用信息技术以实现知识的俘获、存储和重用^[1~3]。KM的基本目的是保存过去俘获的知识,使适当的人员在适当的时机能快捷地

查询到以适当样式表示的所需知识,帮助解决工作实践中遇到的问题。

XML提供了强大的与平台和编程语言无关的数据封装能力,可将应用程序的商业逻辑同数据进行分离;并且作为一种元标记语言,XML在知识管理中可发挥重要的作用。笔者首先提出了一种基于本体论的知识管理结构模型,之后详细讨论了如何利用XML技术来实现知识的俘获、存储和重用。

2 本体的定义

本体概念被引入人工智能、知识工程等领域后被赋予了新的含义。然而不同的专家学者对本体的理解不同,所给出的定义也有所差异。

人工智能领域的学者Neches(1991)等人对ontology进行定义,即:本体是构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延的

规则的定义^[1]。Neches 是最早对本体定义进行研究的学者,从内容的角度给出了本体定义,概括出了本体的基本要素,包括领域术语、关系和规则。这为其后各领域学者对本体的定义研究提供了参考借鉴。

中国学者对本体定义也做了很多研究。张晓林教授(2002)认为“ontology”是概念集,是特定领域公认的关于该领域的对象及其关系的概念化表述。中国标准化研究院的李景(2005)博士认为,本体是一个关于某些主题的,层次清晰的规范说明。北京大学的汤艳莉、赖茂生教授认为 ontology 作为语义网的重要组成部分,是对世界或者领域知识、概念、实体及其关系的一种明确的、规范的概念化描述。张秀兰教授通过对国内外各领域本体定义的深入研究,总结出了本体定义:本体是通过描述、捕获领域知识,确定领域内共同认可的概念和概念间的关系,以用于领域内的不同主体之间交流与知识共享的形式化规范说。

对比中外学者关于本体的定义不难发现,国外对本体的研究较为深入,所给出的本体定义比较经典,被学术界广泛引用。而国内关于本体的研究起步较晚,多数学者对本体的研究是在国外学者的基础上进行的,对于本体的定义强调领域概念,是对本体定义的进一步扩展。关于本体定义,学术界一直没有统一的定论,但存在基本的共识,即本体包括:概念化、形式化、可共享、明确、描述领域知识这五大特征。这五大特征基本概括出了本体的实质内容,但随着本体理论和技术的不断拓新和发展以及本体应用的日渐成熟,对于本体的认识会更加清晰,本体定义也将会更加全面准确。

3 基于本体论的知识管理系统结构

知识管理系统最基本的功能是实现组织内的知识的共享,因此,知识的俘获不仅是整个知识管理过程的开端,也是最基本的要

求,知识的俘获就是将系统内的各种信息转换为知识。知识管理的另一个功能是便于知识的重用,为此,系统内的知识必须与人连接成一个有机的人机共栖系统。

要实现上述功能,必须建立一个知识的高效存取结构,对所关注的信息体中的知识做语义清晰的、概念化的形式描述,以指导建立知识项。可以构筑这样一个基于知识的系统,这个系统就是通常所说的组织记忆(OM, Organizational Memory)。可见知识管理系统应该是建立在组织记忆之上的信息存取技术。本体论(Ontology)为信息系统提供了语义清晰、赞同一致的共同理解基础。

对于知识管理系统,本体论技术更是系统对于知识俘获、存储和重用的内在要求:

- 1) 对系统企事业的各种信息体作语义清晰的说明,可以将非结构化、半结构化的信息体转化为含结构化存储的知识项;

- 2) 本体论的层次结构便于数据在关系型数据库中的存储;

- 3) 本体论提供的术语使精确检索信息成为可能,有望解决当今基于全文匹配检索造成的“信息过载”问题。

综上所述,笔者认为知识管理应以组织记忆为基础,以本体论为线索,以 XML 为表现形式系统。并提出了如下的知识管理体系结构(见图1)。

如图1所示,基于本体论的知识管理体系结构分为四层,由下到上,依次为:

- (1) 信息来源层。包括非结构化或半结构化的文本和其他信息,它是系统信息的提供者,如数据库、企业人员信息、电子邮件和文档信息等。

- (2) 组织记忆层。该层表示知识仓库,存储结构化的知识。借助本体论,将信息体加上元知识(来自于本体论),经标注器处理后,变成为结构化的知识项,最终表示成为 XML 文档,存储于 KM 知识库中。

- (3) 知识服务层。该层提供了知识的服务引擎,包括信息推送引擎、信息检索引擎和信息过滤引擎。其中信息推送指依据被服务者的个人爱好和信息订制,系统及时向其主动提供知识。这些服务引擎依据特定的

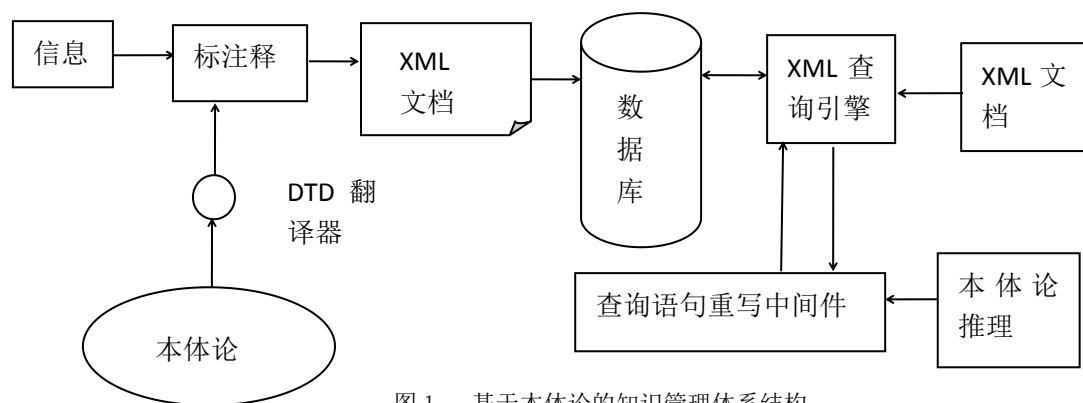


图1 基于本体论的知识管理体系结构

协议和协作模型,负责不同应用程序数据请求格式(不一定是XML形式)的信息发布,同时既要通过必要的策略保证数据的一致性,又要实现必要的访问权限控制。

(4) 应用层,即用户界面层。根据对知识系统的具体应用需求和用户计算环境,利用系统提供的服务引擎,开发合适的应用。

4 系统技术实现

在企事业的信息系统中,存在大量的结构化、半结构化的知识,还有许多非结构化的信息,为了将这些资源结合进知识管理系统,以及重用这些知识,提出了基于XML的知识俘获与检索结构。

4.1 XML 技术简介

Extensible Markup Language (可扩展标记语言), 简称为XML, 是一项针对网络应用的面向内容的新技术, 它具有更多的结构和更多的语义, 良好的可扩展性、自描述性, 简单而易于掌握等特点。它使得原先杂乱而不可扩展的信息交换变得灵活而又规范起来, 使得不同应用之间的信息交换达到了一个前所未有的方便程度。

XML的优点在于将用户接口和结构化资料相分离, 允许不同来源的资料无缝集成, 并可以对同一资料进行多种处理, 既满足了不同用户的需求, 又保证了资料的安全性。

从资料描述语言的角度来看, XML是灵

活的、可扩展的、有良好结构和约束的, 具有更强大的描述能力; 从数据处理的角度来看, XML足够简单并易于阅读, 同时又易于被应用程序处理。DTD(Document Type Definition)是用来为XML文件提供数据结构的文件。

DTD^[4,5]规定了一套关于标记符号的语法、语义规则, 比较准确地描述文本的内容、含义、结构、特征和关系等信息(称为元资料), 而把资料的表现形式交给样式表(XSL)处理, 这样就把资料的内容和表现形式合理地隔离开, 从而大大提高XML资料的可理解性、可交换性和重用性。

基于XML的上述特性, 考虑把XML作为知识管理系统中知识存储的形式和重用的基础。

4.2 知识的俘获存储

在知识的俘获阶段, 首要的任务是要将系统信息资源结合进知识管理系统中。如图2所示, 借助本体论, 将信息体加上元知识(来自于本体论), 经标注器处理后, 变成成为结构化的知识项。标注器的功能就是将关于信息体的元知识, 如内容的表示方式和结构、内容的结构化简介, 以及信息体建立时的上下文(建立和修改时间、作者及身份)等, 转化为OM中的知识项, 标注以后的内容序列化(serialization)为XML文件后, 存储于KM数据库中。

元知识的获得经过了从本体论到DTD的翻译。下面给出把本体论翻译成DTD的四个部分。

1) 对于具有 is-a 关系的概念, 翻译为实体引用语句;

2) 本体论中的概念及属性名翻译为 DTD 中的子元素, 如 Concept person;

3) 本体论中的概念及属性名也可以翻译为 DTD 中的元素的属性, 即 ATTLIST 部分;

4) 从本体论属性产生的元素属性被定义为子元素, 它们可以是原子值, 如 #PCDATA, 也可以是引用元素如 %Person 等。

4.3 知识的检索

检索存储于 KM 数据库中的数据是知识重用的一个重要方面。图 2 中给出了知识检索的过程 (其中 XML 查询引擎采用 XQL^[6] 查询语言), 它包括:

1) 写信息检索组件的界面, 形成 XQL 查询语句;

2) 中间件接收到 XQL 询语句, 并分析找到本体论

中定义术语;

3) 中间件调用本体论推理重新改写 XQL 语句;

4) 改写以后的语句送给 XQL 查询引擎, 从数据库中查找并返回结果;

5) 返回的结果经过中间件的拼装, 送回信息检索组件的界面呈现给用户。

5 基于本体的知识管理的应用

目前国内虽对基于本体的知识管理的实践应用进行了全面深入的研讨, 也取得了一定的研究成果。如何针对具体领域具体需求构建合适地本体, 有效地嵌入到该领域的知识管理系统, 如何运用构建的领域本体快速准确地找到该领域隐形的具有重要价值的知识, 从而实现知识的重用、再创造等, 如何设计并实现适合具体需求的基于本体的知识管理系统将成为未来该领域实践应用的研究趋势。

6 结语

知识管理的作用是便于知识的俘获、存储和重用, 围绕这个中心, 以本体论为线索, 可以设计一个实用的 KM 系统, 同时它也是实现语义 Web 的基础。

参考文献:

- [1] Staab S, studer R, Schnurr H-P, et al. Knowledge Processes and Ontologies[J]. IEEE Intelligent System, 2001, 16(1): 26-34.
- [2] Preece A, Flett A, Sleeman D, et al. Better Knowledge Mangement through Knowledge Engineering[J]. IEEE Intelligent System, 2001, 16(1): 36-43.
- [3] Fischer g, Otswald J. Knowledge Mangement Problems, Promise, Realities, and Challenges[J]. IEEE Intelligent System, 2001, 16(1): 60-72.
- [4] Fallside D C (IBM). XML Schema Part 0: Primer[Z]. W3C Recommendation, 2001-05-02.
- [5] Brion PV. XML Schema Part 2: Datatypes[Z]. W3C Recommendation, 2001-05-02.
- [6] W3C. XQL user GUIDE[EB/OL]. http://www.cuesoft.com/docs/cuexsl_activex/xql.users_guide.htm, 2000-03.