

Ontolingua Server: 全球第一个本体服务器

田晓迪^{1 2}

¹(中国科学院文献情报中心 北京 100080) ²(中国科学院研究生院 北京 100049)

【摘要】 作为全球第一个本体服务器, Ontolingua Server有着较好的构建机制。本文详细介绍了 Ontolingua Server的构建、维护和应用, 以及 Ontolingua Server本体库的本体复用机制, 分析了 Ontolingua的可借鉴点, 并与当今的一些本体系统做了对比分析。

【关键词】 Ontolingua Server 本体 构建 创建 复用 **【分类号】** TP311

Ontolingua Server: The First Ontology Server in the World

Tian Xiaodi^{1 2}

¹ (Library of Chinese Academy of Sciences Beijing 100080, China)

² (Graduate University of Chinese Academy of Sciences Beijing 100049, China)

【Abstract】 As the first Ontology Server in the world, Ontolingua Server possesses good architecture and mechanism. This paper introduces the architecture, establishment and application of Ontolingua Server, as well as the mechanism of Ontology reuse. Besides, the advantages of Ontolingua Server are analyzed and compared with the current Ontology systems.

【Keywords】 Ontolingua Server Ontology Architecture Establishment Reuse

1 Ontolingua Server概况^[2 4 6 8]

Ontolingua Server是由斯坦福(Stanford)大学知识系统实验室于1995年创建的, 这个项目由ARPA和NASA Ames研究中心提供资金支持, 现在通过互联网供全球使用, 用户只要登录 <http://Ontolingua.stanford.edu/> 按照 Ontolingua Server的要求进行注册后即可使用。

Ontolingua Server的设计人性化, 用户不需预先学习; 它支持分布式用户合作建设本体, 用户可通过网络为基础的本体编辑器建设本体; 在信息整合、知识的互操作以及知识库的研究领域中, 本体的复用日益重要。Ontolingua Server具有一个资源丰富的本体库, 用户可以通过整合本体库中的若干模块来形成一个新的本体; 另外, Ontolingua项目开发了一系列工具和服务使得物理上分散的组织在使用本体时能互相协调, 用户也可以通过这些工具在 Ontolingua Server上发布、浏览、创建和编辑本体; Ontolingua Server除了可以建设本体外也被应用于各种工程中。H-Helper就是它在医学上的成功应用, H-Helper是针对患有 HIV 病人的一个计算机门诊病人档案系

统。为了确定病人是否可以参加临床试验, H-Helper采用了一个药物类型和特效药物的本体, 借助 Ontolingua Server帮助决策病人是否符合药物试验的标准。

2 Ontolingua Server

2.1 Ontolingua Server的构建体系^[8 11]

Ontolingua Server提供了一系列的工具和服务来促进本体的使用和知识层代理的交互:

- (1) 一个半形式化的表示语言。Ontolingua Server使用一种扩展版本的 Ontolingua 语言, 它除了具备知识交换格式(Knowledge Interchange Format KIF)详细说明的完全一阶逻辑外, 还提供了便于建设框架的语法。
- (2) 可从知识库中浏览和检索本体。本体的浏览需要以一种简单易懂的格式来表示形式化描述, Ontolingua Server做到了知识的表示和浏览都简单易懂, 符合以对象为导向的结构。
- (3) 在知识库中可以整合、定制和扩展本体。这需要服务器能够识别和解决整合本体中的命名冲突问题, 增加相关词条描述, Ontolingua Server对语言进行了扩展, 所以用户可以从本体库中整合出新本体。
- (4) 方便的本体主题访问保证了用户可以获得最新的词条定义, Ontolingua Server定义了网络协议和应用程序界面。

收稿日期: 2005-08-29
收修改稿日期: 2005-09-09

利用网络开发环境支持分散用户合作开发本体。

Ontolingua服务器提供了一个本体库,允许用户利用库中的本体创建新的本体,同时也可以改进原有的本体,与 Ontolingua Server进行交互主要有三种方式。

(1)分散的组织通过 Web浏览器浏览 Ontolingua Server创建和维护本体。服务器通过 HTTP协议和 HTML和用户进行交互,大量用户可以同时访问 Ontolingua Server。任何熟悉一般网络浏览的用户都可以在 Ontolingua Server上浏览,创建和维护本体。同时,服务器还允许多用户同时在本体上工作;Ontolingua的编辑器通过一系列功能来支持用户合作建设本体。

(2)远程用户有时会需要查询和修改服务器上的本体。Ontolingua Server通过使用应用编程接口(API)扩展了框架协议,所以它可以通过一个界面支持用户的查询和更新行为。

(3)用户可以为特定的应用而将本体转换成特定的格式。如 KIF格式可以转换为逻辑推理器可识别的逻辑语言文件,CommerceNet通过推论来响应客户机的查询并正确向其发送查询结果。

2.2 Ontolingua语言^[9-11]

最初的 Ontolingua语言是用来支持设计和说明一个逻辑语义明确的本体,为了达到这个目的,KIF诞生。KIF是一个具有简单语义的一阶逻辑语言,即形式化语言,它主要用于知识表示形式不同的计算机程序之间的知识交换。KIF的3个基本的概念类为对象、函数和关系,提供了函数和关系定义操作符,支持函数和关系的属性表示,支持元知识的表示。

Ontolingua将 KIF作为它的核心语言,同时也扩展了原有语言和本体开发的便捷性,第一,它支持用户在一个新本体中创建可被整合、扩展和改进的本体模块;第二,它分离了本体的内在描述和外在表现形式,为 Ontolingua语言开发了一个以对象为导向的外部表达方式,本体的内部描述表示为一系列的 KIF公理和非逻辑符号。Ontolingua包含 KIF分析器、本体分析工具和一组 Ontolingua转换器,这些 Ontolingua转换器可将由 Ontolingua语言表示的本体转换为 CML-RE, DL-RE, CLIPS-SENTENCES, CLIPS, PROLOG-SYNTAX, LOOM, EPKIF, GENERIC-FRAME或 KIF等目标语言。所以说,Ontolingua是独立于特定表示系统的本体定义机制。

2.3 Ontolingua Server的开发环境^[5-10]

为了满足用户的开发需要,Ontolingua Server提供了一个可以辅助用户进行浏览,创建,维护,共享和使用本体的环境。

(1)浏览本体

Ontolingua的浏览环境有一个基本功能是从本

体的一个术语直接链接到另一个术语。查看一个本体时,用户往往希望能从一些术语定义中推断出某个术语的信息,Ontolingua的基础框架系统提供了有限的推理集合,包括下位类对上位类的继承、实例等,直接声明以罗马字体出现,而推论信息以斜体字出现,用户需要查看推论信息的原声明时可点击每条推论信息的“GO”按钮。

Ontolingua Server提供了上位类、下位类的浏览方式向用户展示了整个本体的层级。区别以往的节点或图解表示,Ontolingua Server的本体层级通过上下交错的文字来展示,文字旁边的按钮控制下位树的打开或关闭。

(2)创建本体

Ontolingua Server设计了一个与浏览环境外观相似的编辑环境,不同点是编辑环境中有两种新按钮:编辑笔和插入。编辑笔一般出现在可修改的信息前,用户希望改变信息时可选择编辑笔及适合的模式。插入按钮出现在允许添加信息的地方,如对一个属性添加新值。

①维护本体

用户可比较两个本体,查看需要转换的功能设置,监察本体的变化或遗漏的更改;可分析本体存在的矛盾和模糊的术语,检查所有的属性、属性值以确定各种限制,如,领域,范围,属性值类型等;也可将一个本体分成若干本体模块。

非正式文件也是维护本体的重要部分,Ontolingua Server支持非正式文件中的某些特殊关键词作为注解,用户可将一些关键词如 example-verified-by, modification-by和 formerly-named加入到非正式文件中提高本体的结构性。

②共享本体

Ontolingua Server支持本体复用的首要机制就是本体库。当用户认为他所创建的本体可以被他人使用时可在 Ontolingua Server上发布这个本体。这个本体被 Ontolingua Server认可后,其他用户就可以访问 Ontolingua Server的本体库使用它。

③合作开发本体

Ontolingua Server支持分布式合作建设本体:一是 Ontolingua Server对用户和组织的访问进行控制,二是 Ontolingua Server给予多用户的 Session。

Ontolingua Server通过 Group Session支持合作工作。当用户打开一个 Session并将它指定为一个组所有,那么这个组的其他成员都可以加入此 Session并在某个本体上同时工作;如果一个成员对设置做了某些改变,“通告机制”会及时通知组内其他成员;通告可以直接链接到修改后的定义,并且根据添加、删除和修改等操作描述改变。

④易用简单

Ontolingua Server在设计界面时既考虑到让初学者易于掌握,也要有足够的力量支持有经验的用户,为了协调和适应多水平的用户,Ontolingua Server添加了大量的用户参数协控用户的界面行为。Ontolingua Server提供了四种页面:本体库的内容表格、本体摘要页面、结构页面(对类、关系或例子)和类浏览器。超链接设置提供给用户及时的指南、在线文件和上下文帮助。另外,Ontolingua Server还允许用户继续建设已经保存的本体或对其进行多次修改。

(3)使用本体

利用 Ontolingua Server 复用本体的一个方法就是将这个本体转化为另一个系统的表示语言,如 CLIPS LOOM, Prolog。现在, Ontolingua Server 可将本体转换为十种不同的表示,用户可以通过 E-mail 传递相关的转换,也可以用 E-mail 传递本体的来源码或形式化正文等。

2.4 Ontolingua Server 的本体复用机制^[9-10]

Ontolingua Server 具备清晰的形式化语义并允许用户通过包含、多态、完善和限制四种方式从一个模块结构的本体库中复用其已有本体。

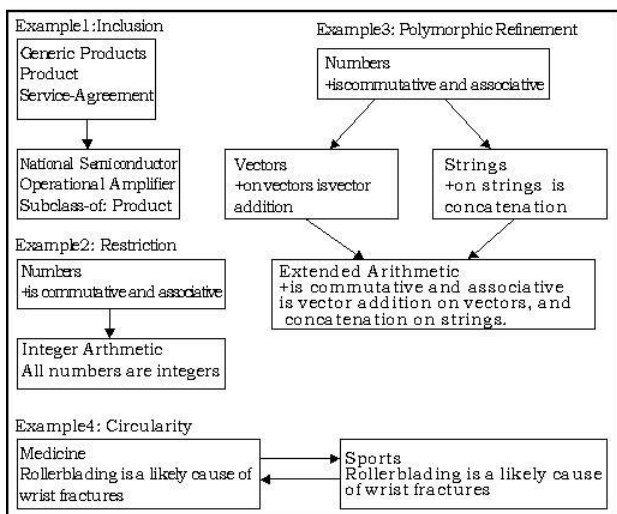


图1 本体建设的例子

图1中显示了本体建设中的例子。例1展示了本体间最简单的关系:包含。一个关于半导体产品的本体开发者需要表示产品、价格、服务等信息,他通过直接复用(包含)本体库中的 Generic Product 本体就达到目的;例2中,我们看到专用本体可能会约束所包含的公理。比如,在 Integer-Arithmetic 本体中,所有的数字都限定是整数;在例3中,作者希望用两种不同的方式来扩展操作符“+”,本体库中包含了关于“加”操作符公理的 KIF-Numbers 本体,作者希望在一个本体中将“加”操作符应用为向量,而在另一个本体中将其应用为串,这种操作称为多形态的完善;在例4中,我们看到本体间的包含循环关系,这两个本体一个是医药方面的,一个是运动方面的。医药本体需要涉及运动本体中的很多术语(如 Roller-blading 是青少年手关节骨折的重要原因),运动本体也必须涉及到医药术语(如举重运动员可能会使用 Anabolic Steroids 加强肌肉的增长)。本体设计者不愿让任何本体被另一个本体的非逻辑符号影响,所以 Ontolingua Server 对这个问题的解决措施主要有两个方面:在内部表示中加入一个本体包含操作符;重命名符号和消除来源于

多组件本体的非逻辑符号的歧义。

(1)在表示形式中加入本体内含物

Ontolingua Server 提供了一个本体可包含另一个本体的功能。每个本体都是由非逻辑符号的词表和一系列公理说明的。在本体 B 中要包含本体 A 需要详细说明 A 词表到 B 词表的转换,即将 A 中的公理进行转换然后加入到 B 的公理中,形成的公理集称为“本体 B 的公理”。因此,如果 B 被包括在本体 C 中,那么 C 就包括 A 和 B 的转换公理。所以,“本体 O 的公理”就是本体 O 自己的公理和 O 中其他本体转换后的公理。这种包含的思想定义了可有循环包含关系的图表。Ontolingua Server 允许本体包含是及物的,即如果从 A 到 B 的本体包含图表中有一条路线,我们可以说本体 B 包含本体 A。

Ontolingua Server 通过分离每个本体的符号词表消除了本体在内部表示中的符号冲突。在内部表示中,本体 A 词表中的符号 N 和本体 B 中的符号 N 是不同的,即每个本体都为自己的符号提供了一个本地命名空间。我们可以注意到,因为本体包含就是公理集的联合,所以循环的包含图表就不是一个问题了。

Ontolingua Server 允许使用者声明本体间的直接包含关系和基于公理的符号创建包含关系。也就是说,如果本体 A 包含了一个公理,这个公理涉及到本体 B 的符号,那么这个系统就认为 B 包含在 A 中,这个包含规则说明被引用的符号是本体中的一部分。

(2)解决组合本体的符号问题

Ontolingua Server 所增加的包含操作为本体的组合和复用提供了一个明确有效的方式。然而为了消除歧义,符号有时候需要增加扩展名来保证唯一性,但用户可能不熟悉这个名字,此外,用户不能执行一些操作,如重命名被包含本体中的符号或选择性的引进被包括本体的符号或输出它的符号到其他的本体。Ontolingua 通过将符号转化为文字输入到内部符号表示中(从内部符号表示中将符号转化然后输出)解决了这些问题。

非逻辑符号一般会在某个本体中定义并有一个唯一的名字,定义这个符号的本体就称为此符号的“Home Ontology”。Ontolingua Server 根据所给本体的角度说明输入输出的符号,比如,如果一个名为“N”的符号是在本体 A 中定义的,而另一个也叫“N”的不同符号是在 B 中定义的,所以从 A 的角度来看,“N”说明为“在 A 中定义的名为 N 的符号”,而从 B 的角度看“N”说明为“在 B 中定义的名为 N 的符号”。这是一种默认的观点,Ontolingua 编辑器或本体来源文件都可以准确的解释其意思。另外也可在本体的名字后加一个后缀字符“@”命名这个符号,比如说,从本体 A 的角度看,名为“N”的符号可以写为 N@A 这样命名的符号叫做“Fully Qualified Reference”,Fully Qualified Reference 使定义在任何本体中的符号都能被其他本体所引用。

当从需要角度看符号名本身已没有歧义时, 后缀名就不需要了, 这种符号叫做“Recognized By name”, 一个 Recognized By name的符号一般是从其 Home Ontology 的角度来看的, 所以决定一个符号在一个非 Home Ontology 中能否 Recognized by Name 还需要其他的规则。Ontolingua 允许用户重命名本体中的符号, 然后从这个本体的角度看此符号就是 Recognized by Name。重命名的规则包括本体名、一个符号参考、一个将在此本体中使用的本地符号名, 这样 Ontolingua 在输入输出时就可以识别相同符号的不同名字, 开发者可用适当的名字引用其他本体的符号, 多本体间的符号命名冲突也就得以解决。

我们利用图 1 中的例子看看 Ontolingua Server 是如何支持本体包含、循环从属和多态性完善的。例 1 中的本体包含关系, 通过 Ontolingua Generic Products 本体中的符号如 Service-Agreement 并不和新本体中的符号名冲突, 所以系统将其默认的输入到新本体中, 因此从原本体角度看它的名字还是 Service-Agreement 不用加后缀; 例 4 中的循环从属可以通过 Full Qualified names 来引用其他本体的符号, 继而可以解决所存在的问题。比如说, 在医药本体中, 涉及到 Roller-Blading 就用 Roller-Blading@Sports 表示, 这样医药本体包含符号 Roller-Blading@Sports 的相关公理但不会同时包括运动本体的其他符号; 例 3 多态性完善中的“+”, 如果本体 x 不包括 Vectors Strings 或 Numbers 本体, 那么在 x 中涉及到 +@Vectors and +@Strings 将使 x 包括 Numbers 但不包含 Vectors 或 Strings 这是因为 +@Vectors and +@Strings 所涉及的符号其 Home Ontology 是 Numbers 如果 Vector 本体的 Vector 增加公理或 String 本体的 String 增加公理被包括在 X 中, 那么用户必须明确的声明包含关系。

3 Ontolingua Server 在实际工程中的应用^[4-7]

3.1 CommerceNet

斯坦福大学、合成技术中心 (Center for Integration Technology CII) 和知识系统实验室 (Knowledge Systems Laboratory KSL) 参与了由 ARPA 发起的 CommerceNet 工程。CII 和 KSL 的总目标就是展示互联网上产品和数据的目录功效, 用户可以通过具体或规范的描述 (不仅仅只是关键词或号码) 来查找所需的产品信息, 还可以跨目录进行产品比较。这个工程的一个目标就是开发一个具备产品分类的术语和关系描述的本体, KSL 本体编辑器是主要的创建工具。

3.2 企业工程

人工智能应用协会 (Artificial Intelligence Applications Institute AIAI) 的企业工程的目标是提供方法和计算机工具集合来获得一笔交易的相关信息并对其分析, 然后鉴别

和比较能够达到交易要求的选择。合成方法和工具的框架基于一个适合企业模型的本体, 它支持应用于商业的以实际知识为基础的模型工具和方法。企业本体是一个和商业企业相关的术语和定义的集合, 它提供了一个通用语言来帮助商业软件的应用和商业交流。因为企业本体在企业工具集中是非常重要的 (软件应用间的交流语言是 KIF), 所以它必须是形式化的。项目组首先用自然语言来开发本体, 然后使用本体编辑器将其转化为 Ontolingua

3.3 InteMed 工程

InteMed 工程在医药信息学研究中是一个合作型的工程, 它由五个不同的研究型站点共同合作: Columbia 哈佛大学 (决策支持组)、斯坦福大学的 Mass General Hospital (计算机科学实验室)。斯坦福组在试图建设一个医学词汇服务器时最大程度上使用了 Ontolingua 服务器, Vocabulary 主页上有 InteMed 组已建设的医学本体的列表。

3.4 试验库

众所周知, 世界对临床试验都作出了很大的投入, 但是试验结果却很难解释。这个工程的目标就是利用 Ontolingua 来形成一个有关临床试验领域的形式化规范和实现临床试验数据库中的知识共享。

3.5 基于网络的信息代理

这个工程的目的是通过一些关键性技术使买主和卖主可以创建和维护基于网络的信息代理, 代理可以从网络上各种卖主目录和数据库中检索服务和产品信息。以一个及时有效的方式获得相关信息的能力是代理的关键, 基于计算机的信息代理将为大型团体提供这种能力。信息代理的创建及维护技术的普遍适用性将促使以网络代理服务为主要产品的新型工业的建立。

4 Ontolingua Server 的可借鉴点

4.1 Ontolingua Server 的优点^[1, 3, 12]

(1) Ontolingua Server 是非常可靠的。从它的公共发布日期 1995 年 2 月起到现在, 它在 99.89% 的时间都是可访问和使用的。Server 的不间断服务可以维持两个星期, 而这些可靠服务都是促使远程用户使用其工具和服务的基本原因。

(2) Ontolingua Server 已经达到了大量用户持续使用和访问的水平, 它为各种水平的用户提供了合适的服务, 满足了他们科研和学习的要求。

(3) 界面简单易懂。对工具 and 软件的使用说明细致明确, 每个页面都有帮助按钮, 用户可以及时查询。

4.2 Ontolingua Server 与其他本体系统的优势比较^[13]

与菲利普·马丁 (Philippe Martin) 开发的 WebKB-2

相比, Ontolingua Server是基于模块的本地服务器,用户可以添加、修改或复用本体模块,促进新本体的建设;采用以对象为导向的设计思想,用户不必一一了解原有本体的内部设计,选中具备所需功能的本体就可直接复用,而 WebKB-2是以类为单位,用户在添加或复用类时还必须要了解其相关的上位类、下位类等,所以 WebKB-2的复用功能明显弱于 Ontolingua Server。

与 Guarino于 1999年开始研发的 Ontoseek信息检索系统相比, Ontoseek 着重于精确地或详细地描述黄页中的产品或服务,将一个本体驱动的内容匹配机制与一个具有中等表达能力的形式化表示系统相结合,尝试如何与本体和大辞典数据库相集成,为用户提供一个可以使用领域内任何词汇进行交互式语义查询的系统,而 Ontolingua Server则是侧重于通过本体的浏览、编辑、修改和复用来支持用户合作开发新的本体,两者的目的不同,因而结构功能也有各自的特色。

与美国普林斯顿大学 (Princeton University)认知科学实验室开发的 WordNet相比, WordNet是基于心理语言学规则的英文辞典,它以同义词集 (Synests) (在特定的上下文环境中可互换的同义词集合)为单位组织信息,对查询结果的演绎比较符合人类的思维定式,但是 WordNet的实质只是一部电子词汇的数据库,它不具备推理功能,而 Ontolingua Server已开发出专用的推理引擎,在本文第一部分所提到的 H-Helper就是它在医学上对于推理功能的成功应用。

5 结 语

Ontolingua Server和其他本体开发环境的明显区别在于它的支持分布式合作建设本体和拥有一个“可复用的本体海量存储库”。Ontolingua Server支持多用户同时进行操作,用户可以通过使用以网络为基础的本体编辑器来建设本体,可以通过组合和扩展在线本体库的本体来创建新本体,大大的促进了本体的利用和新本体的创建,目前 Ontolingua本体库中大约包含上百个本体,每个本体都有相应的介绍,用户可以根据需要选择适合所建本体层级的某个本体,它能结合和协调不同本体的公理、定义和非逻辑符号,为本体的复用扫除了障碍;它支持循环的包含图表,用户能用多个方式扩展本体;包含了很多默认的设置,使开发

者能专注于较为细节和复杂的控制,节省了精力和时间。

参考文献:

- 1 Ontolingua frame-editor <http://www-ksl-svc.stanford.edu/5915/doc/frame-editor/> (Accessed May 14 2005)
- 2 ServerHelp <http://www-ksl-svc.stanford.edu/5915/doc/server-help.htm> (Accessed May 15 2005)
- 3 Knowledge Sharing Effort public library. <http://ksl-web.stanford.edu/knowledge-sharing/> (Accessed May 14 2005)
- 4 KSL Ontology Server Projects <http://www-ksl-svc.stanford.edu/5915/doc/Ontology-server-projects.htm> (Accessed May 15 2005)
- 5 A Guided Tour to Developing Ontologies Using Ontolingua <http://www-ksl-svc.stanford.edu/5915/doc/frame-editor/guided-tour/index.html> (Accessed May 16 2005)
- 6 Ontolingua <http://www.ksl.stanford.edu/software/Ontolingua/> (Accessed May 16 2005)
- 7 Relevant papers <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 16 2005)
- 8 Fikes R., Farquhar A., & Rice J. Tools for Assembling Modular Ontologies in Ontolingua Knowledge Systems Laboratory April 1997. <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 20 2005)
- 9 Fikes R. & Farquhar A. Large-Scale Repositories of Highly Expressive Reusable Knowledge Knowledge Systems Laboratory April 1997. <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 20 2005)
- 10 Farquhar A.; Fikes R.; & Rice J. The Ontolingua Server: A Tool for Collaborative Ontology Construction Knowledge Systems Laboratory September 1996 <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 23 2005)
- 11 A. Farquhar R. Fikes W. Pratt & J. Rice Collaborative Ontology Construction for Information Integration Knowledge Systems Laboratory Department of Computer Science KSL-95-63 August 1995. <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 23 2005)
- 12 Chaudhri V. K.; Farquhar A.; Fikes R.; Kap P. D.; & Rice J. P. Open Knowledge Base Connectivity 2.0 Knowledge Systems Laboratory January 1998 <http://www-ksl-svc.stanford.edu/5915/doc/project-papers.htm> (Accessed May 23 2005)
- 13 李景. 本体理论及在农业文献检索系统中的应用研究——以花卉学本体建模为例. [学位论文]. 北京: 中国科学院文献情报中心, 2004

(作者 E-mail: tianxd@mail.ksl.ac.cn)

《现代图书情报技术》2006年(年刊)征订通知

《现代图书情报技术》2006年(年刊)即将出版。每册定价 30元(含邮费)。内容涉及数字图书馆技术、图书馆自动化、信息检索技术、网络资源与建设、网络多媒体技术、图书情报技术工作交流等方面。如欲订购,请直接将款汇到:北京中关村北四环西路 33号《现代图书情报技术》编辑部 邮编: 100080 本刊编辑部收,款到后即将发票随刊一并寄出。印数有限,欲购从速!

编辑部联系电话: (010) 82624938

(编辑部)