

# 知识库系统中语义网知识的表示

吴 丹<sup>1</sup>, 易 辉<sup>2</sup>

(1. 华中师范大学信息管理系, 武汉 430079;

2. 华中科技大学电子与信息工程系, 武汉 430076)

**摘 要:** 基于对知识库系统中知识表示的理解, 提出了在语义网环境下知识库系统的建立和相应的知识表示, 并将此扩展到语义网中整个的知识表示系统。

**关键词:** 知识库系统; 语义网; 知识表示

**Abstract** Based on the understanding of knowledge representation in knowledge base system, the author put forward the construction of knowledge base system and knowledge representation for semantic Web. Moreover, it is expanded to the whole knowledge representation system.

**Key words** knowledge base system; semantic Web; knowledge representation

互联网的发展现已进入关键的转折期, 这一巨大信息资源中所孕育的无穷潜力促使着人们积极探索它的未来。虽然数字鸿沟和关键技术上的难题使我们对其最后的结果还需要等待, 但是按需服务与无处不在的计算作为互联网的发展方向却从未被改变过。站在下一代互联网的边缘, 被视为演进方向的语义网正是这种发展延伸的一个侧面。为此, 我们有必要对此进行一定的分析和研究。

## 1 知识库系统、知识表示与语义网

### 1.1 知识库系统

知识库是以描述型方法来存贮和管理知识的机构, 由知识和知识处理机构组成, 形成一个知识域。该知识域中除了事实、规则和概念之外还包含推理、归纳、演绎等知识处理方法, 逻辑查询语言、语义查询优化和人机交互界面等。知识库系统是利用计算机所存贮的知识对输入的数据进行解释, 生成作业假说并有对其进行验证功能的一个系统, 其核心组成部分是以一致的形式存贮知识的知识库

和使用知识库藏内的知识执行推理的控制机构。与数据库系统不同, 知识库系统根据输入的数据信息利用知识库在自身进行推理过程后, 提供给系统使用者的是判断分析后的结果, 而不仅仅是对用户提供可检索的信息。知识表示、知识利用和知识获取是知识库系统实现的三个关键技术。

### 1.2 知识表示

知识是人类改造现实世界的实践活动的认识和经验的总和。表示是为描述世界所作的一组约定, 是将知识编码成一种适当的数据结构的符号化过程。知识表示则是寻找知识与表示之间的映射, 选择合适的形式来表示知识, 即将关于世界的事实、规则、概念等编码成为一种合适的数据结构, 以计算机所能接受的形式表示出来。它研究的问题是设计各种数据结构, 以知识的形式表示方法; 研究表示与控制的关系; 表示与推理的关系以及知识表示与其他领域的关系。目前主要的知识表示方法有: 逻辑表示模式、语义网、过程表示和产生式系统、特性表、框架和脚本。

收稿日期: 2003-11-06

作者简介: 吴 丹 (1978-), 女, 硕士研究生, 研究方向: 知识组织与数据挖掘; 易 辉 (1977-), 男, 助教, 研究方向: 网络信息管理。

同一知识可以采取不同的表示方法,但在解决某一问题时,不同的表示方法可能会产生完全不同的效果。因此,为了有效地解决问题,我们必须选择一种合适的表示方法。

### 1.3 语义网

W3C组织在对现行互联网进行扩展时提出了语义网(The Semantic Web)的概念,并试图通过提供有意义的网页结构来建立一种通过软件实现网页间因复杂工作而链接的环境。这种设计使得整个互联网可以自动处理,所有内容易于使用,构建了一个遍布全球、可理解的平台,允许知识自动共享和处理。在语义网中,计算机将根据关键名称定义的超链接和逻辑推理规则发现语义数据的含义,其最终结果就是能够刺激开发自动化的网络服务。语义网的目标是扩展当前的

WWW,把信息表示为计算机能够理解和处理的形式,即带有语义,便于人和计算机之间的交互与合作,又能增进资源的共享。智能信息处理是研究的一个侧重点。

## 2 知识库系统中的语义网知识表示

### 2.1 知识库系统的知识表示

知识库系统的实现需要解决知识表示、知识利用和知识获取三个关键技术。知识表示即知识表示的形式,计算机要能够进行处理,并以一种人类能理解的方式告知处理结果。知识利用是指利用知识库中的知识进行推理,从而得出结论的过程。知识获取是指从知识源获得知识来建造知识库的工作,如图1所示:

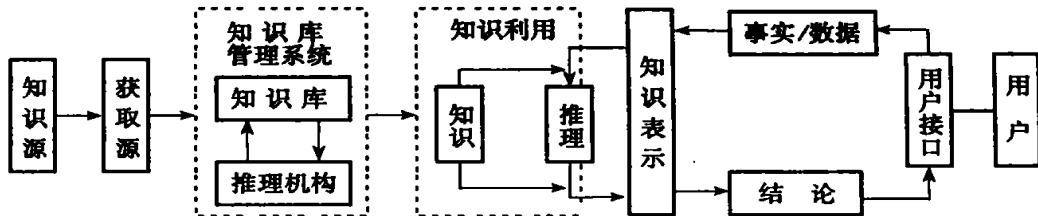


图1 知识库系统

在计算机科学中,一个好的解决办法往往依赖于一种好的表示形式。计算机可以处理的问题必须可用形式化表示,针对问题必须存在一个算法,且这个算法必须是可计算的。在将计算机应用于数值计算时,原问题和形式化后的问题是一致的,也就是说数据结构和算法的语义是直接的,其间没有转义映射。在将计算机应用于知识处理时,由于面对形形色色的问题,原问题和形式化后的问题在理论上是同态的,也就出现了知识表示问题,即为了计算必须将非数值语义和语义间的关系转化为符号,用指称的办法来保持和恢复计算结果的语义。在已开发面世的系统中,知识表达方式的选择至关重要,它不仅决定了知识应用的形式,也决定了知识处理的效率和可实现的域空间规模的大小。

对于知识库系统而言,要选择或设计一

种合适的表示模式,应考虑以下问题:知识表示模式的表示能力?是否有足够的表达能力以便于应用中对知识的各种运用(例如推理、求解等操作)?知识表示的模块化程度?是否便于知识的扩充和修改?是否支持自顶向下、逐步求精的设计原则?是否具有足够有效和强大的推理能力?表示是否严格、简洁等等。科学家已提出了多种知识表示方式,如:逻辑表示模式、产生式系统、过程表示、框架、特性表、脚本,当然还有语义网。

### 2.2 知识库系统中的语义网知识表示

语义网的新一代网络技术是为专门改善使用不同术语的人所进行的网上交流和提高数据库的互用性而设计的,同时也提供了一种人与网上的多媒体资源交流的工具,使人机间的操作更具有互动性。为了实现语义网的功能,计算机必须做到能够符合信息的结

结构化收集和自动推理的法则。Tim Berners-Lee 给出了语义网中的层次关系,以完成基于语义的知识表示和推理,如图 2 所示:

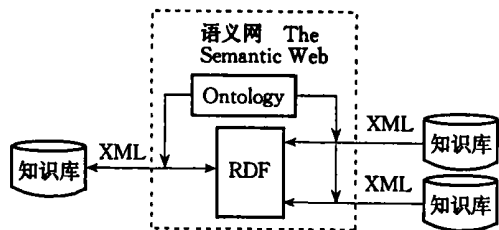


图 2 语义网在知识库系统中的应用

XML (eXtensible Markup Language, 可扩展标记语言) 为计算机提供可分辨的标记, 定义了每一部分数据的内在含义。脚本 (或者说程序) 可以利用这些标签来获得信息。XML 允许使用者在他们的文档中插入任意的结构, 但没有说明这些结构的含意。

资源描述框架 (Resource Description Framework, RDF) 是用来表达含意, 是一种用于支持语义网的语言。RDF 是为互联网开发的一个支持 XML 资源描述的框架或元数据, 提供了一个支持 XML 数据交换的主动宾三元组结构。

Ontology (本体分类) 是概念化的详细说明, 用来定义描述某一类领域知识的术语及相互关系。一个有意义的 XML 代码的网页可以通过由网页到 Ontology 的指针来互动, 从而扩大网页的功能。Ontology 的推理法则提供了更强大的功能, 可以通过简单的样式实现精确地查找。

语义网是一个受限域的智能系统, 不能只简单地在知识库中编码进更多的规则和事实, 这不但困难, 而且也影响性能。要实现语义 Web, 使机器能够处理信息, 首先必须建立更高层次的互操作性标准。这些标准的共同目的在于表示网络甚至网络以外的知识, 属于知识表示的范畴。

知识表示简单地说是程序怎样对现实世界建模, 研究范围应该是知识表示方法, 即用什么样的描述方式最有利于程序的自动处理和自动推理。语义网网络最初是在自然语言

的理解系统中, 为了表达 Network 的意思而设计的一种表示方法。这种方法提倡协同工作, 不同的机器、程序都能从不同的资源中收集网页内容, 处理成可读信息并相互交换结果, 提供自助服务。当数据随语义到达时, 利用通道匹配和继承进行推理, 以保证网络操作所得结论的有效性。由于分布式和开放性的 Web 特征, 语义网条件下的知识表示具有传统知识表示领域所不具备的一些特征, 这些特征也是对知识表示语言的必然要求。在语义网的环境下, 数据描述能力和规则描述能力是知识表示语言最先考虑的问题, 而系统由此带来的不可判定性和二义性却无关紧要。

采用语义网络表示的知识库的特征是利用带标记的有向图来描述可能的世界, 如图 3 所示, 它实际上是对知识的一种图示方法。

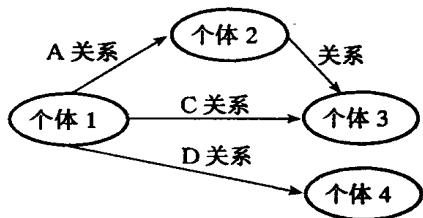


图 3 语义网知识表示

在语义网络中, 个体可以用图的一些节点来表示, 节点之间通过一组带有标记的弧连结, 带标记的弧表达了节点之间的关系。这样语义网就可以将搜索的水平从词上升到概念, 突破了辩识对象的局限, 可以在数据层面进行分析, 并清楚表达了对对象间的关系。结点表示客体的性质、概念、状况或动作, 带标记的边描述了客体间的关系。知识库的修改是通过插入或删除客体极其相关的关系来实现的。摆在语义面前的挑战是, 提供一种通用的 Web 语言, 能同时表达数据以及根据数据进行推理的规则, 这也是评价知识表示是否适合语义网的根本标准。目前已出现不同的软件代理以不同的算法和不同的规则集使用 Web 上的数据, 但其本身具有的一些基本特征在一定程度上会对语义网 (下转第 44 页)

## 4 结论

移动通信领域的欠费是个令人头疼的问题。基于数据仓库技术,我们提出了使用数据挖掘中的决策树分类算法对移动通信的欠费进行分析,并结合讨论了主成份分析的属性归约方法。分析预测的结果一方面可以呈现给决策树,另一方面它可以为建立有效的预警控制系统提供很好的参数。实验结果表明,

(上接第 11 页)

Web 中的知识表示提供一些思路 and 限制,并最终决定语义网能够达到什么样的高度。采用语义网络表示法比较合适的领域大多是根据非常复杂的分类进行推理的领域以及需表示事件状况、性质及动作之间关系的领域。

## 3 语义网知识表示所面临的问题

对知识表示方法的选择反映了知识库系统的开发者对所模拟的行为本质的认识。不是任何一种表示方式都是对客观的真实反映,无论是符号主义还是连接主义都是对知识的一个近似表示。利用语义网知识表示也会存在如下的一些问题:

首先,Web 具有分布式和开放性的特征,缺乏一些集中的控制机制会使它的发展和组织结构缺乏一个非常严格一致的体系。如果以语义网选择一个知识表示,那么这个知识需要有足够的表达能力。并可以得到不同的知识库和规则集合。为满足通用性的要求,势必会导致描述对象无限制的膨胀。此外,对知识的表示应与平台无关,其语义必须有表示网络甚至网络以外的知识的能力。Web 也就是说所有的逻辑数据必须能够被直接或间接表达,但现实世界本来就充满了不可判定命题和不易处理的难题甚至悖论,因此不能指望通过选择来达到消除的目的。即便是对于相同问题也无法苛求他们必须采用统一的词汇来描述,并且可能还会有不同

所建立的欠费分析预测系统有很高的准确率,能够为移动运营商解决和控制欠费提供有力的支持,在经济上具有很高的回报率。

## 参考文献:

- [1] Han J.等.数据挖掘概念与技术.范明,等译.机械工业出版社.
- [2] 史忠植.知识发现.清华大学出版社.
- [3] 中国移动经营分析系统业务规范 2.0.

的观点,甚至无法保证其中的信息是正确的。网络的动态性也会给语义网知识表示带来问题,主要表现在内容的增减和链接不可预测性地改变,这些改变对于力图获取信息的个人或者程序来说都会有至关重要的影响,更何况是对于一个发现特定的信息而无法全部收集所有需要的知识库系统而言。

建立知识库系统的最终目标就是以网络为基础来实现一个巨大的、虚拟的、分布式的知识仓库,因为它能够利用散布在各处的所有知识来帮助使用者解决问题。尽管要完全到达预定目标还有很多问题需要解决,但这些目标正在通过知识的分层、方法库的引入、智能代理的建立以及知识交换标准及协议的完善而逐步得到实现。我们有理由相信,知识库系统与语义网的结合将促成知识化网络新时代的早日到来。

## 参考文献:

- [1] Hefflin J, Hendler J, Luke S, Shoa A. Prototype Language for the Semantic Web. Information Science, 2001.
- [2] 李德毅.知识表示中的不确定性.中国工程科学, 2000(10):
- [3] 何杭军,龚丽霞,陆汝占.语义网知识表示的评价标准.计算机工程, 2002(6):
- [4] 徐天伟,等.由语义网络语言的知识表示到谓词逻辑的转换.云南师范大学学报(自然科学版), 2001(6):
- [5] 刘柏嵩.基于知识的语义网:概念、技术及挑战.中国图书馆学报, 2003(2):