

基于本体系统的知识管理人工智能研究

程少川¹, 程向前², 张朋柱¹

(1. 西安交通大学管理学院, 西安 710049; 2. 西安交通大学计算机教学实验中心, 西安 710049)

摘要:分析了知识管理中的信息组织难题和群体过程非结构化特性的本质意义,介绍了用于解决人工智能中非结构化问题的本体系统方法,并通过支持群体交互知识管理的电子公共大脑本体系统设计的案例,对本体系统方法的概念、构成和应用方法作了具有典型意义的示例性介绍。

关键词:本体系统; 电子公共大脑; 知识管理

中图分类号: C931

文献标识码: A

文章编号: 1007-9807(2003)02-0017-06

1 知识管理的非结构化信息组织难题及其本质意义

知识管理是新世纪知识经济环境中为世界管理学界关注的热点话题。尽管人们不难理解知识管理中可能存在的巨大利益,但是能够将知识管理付诸实施的企业却不是很多。原因大致存在于组织机制和技术成熟度两个方面。组织机制方面的问题关系到组织的权利、契约关系、管理职能等方面;技术方面的原因则主要存在于信息组织领域。由群体参与的持续的、跨时间和地域的信息交互与存储,本身是一个复杂的非结构化问题,计算机专家和知识管理专家们为此开发了多种多样的技术和方法,但技术的复杂度也导致成本的上升,影响了知识管理的真正展开。可以说,知识管理在技术或方法上的主要障碍,是如何对付知识管理过程中出现的复杂的非结构化问题,即如何使得纷繁复杂的信息变成具有明显结构关系的知识的问题。

解决这一问题,首先应该了解群体过程非结构化问题的实质究竟是什么。

当人们说某一群体交互过程具有非结构化特性时,只是说明这一过程所具有的结构不是先验

的,并且由于它的非重复性,在某种程度上是未被人了解的。所以,群体过程的“非结构化”是主观的概念,是人们对于客观过程认知的状态,而不是事物本身的性质。当一个过程发生以后,它的结构从客观上来讲应该是可以描述清楚的,之所以难以被描述清楚,主要是因为这种结构复杂性较高和人们用于认识或描述结构的方法不够便利。群体交互非结构化过程复杂性具体表现在以下两个方面:

时间序列中的随机复杂性 个体进行信息交互时的信息随机性(如:随时间变化的视角,所关心的问题、态度、灵感等),导致过程结构的复杂。群体过程中,每一个个体都可能有不同的视角,并且在信息交互的激发下,观点和态度会随着过程的不断深入而发生变化。同时,在信息交互过程中,不断产生的新信息及其与其他信息的相互关系是不可预知的。这是一种动态的复杂性,它所产生的信息及信息之间的关系丰富多彩。这种复杂性同时也带来了对于群体过程中任意一个时间点的信息结构认知的难度。

任一时间点上的状态复杂性 群体过程中所形成某一时间点的信息结构,包括信息的内容及其丰富多彩的相互关系。这些关系可以包括逻辑

收稿日期:2001-11-26; 修订日期:2002-05-08。

基金项目:国家自然科学基金资助项目(70042014); 加拿大 CCUIPP 基金资助项目。

作者简介:程少川(1965—),男,博士,讲师。

关系、层级关系、情感关系以及由交互过程中的态度形成的关系。信息之间关系的丰富多彩,导致了由各种关系构成的信息结构状态的多样性。具体信息之间的关系可以是“疑问”、“支持”、“反对”、“否定”乃至“否定之否定”等等。

在自然的非结构化过程中业已形成的信息,其结构应该是可以描述的。描述的方法,则必须从信息的关系中去寻找。信息的意义反映在信息与信息(或信息与事物,事物可以看作广义的信息)的关系之中。信息之间的关系明确时,可以认为信息是有结构的。现在的问题是,在建立系统时,如何来描述信息之间的关系?

形式化本体系统方法,提供了一种较为简洁的,用规范的形式化方法解决这一复杂系统问题的途径。

2 本体系统的概念及其构成

本体系统(方法)和本体论(ontology),在英文的表达上是同一个词。本体论在哲学上的定义是对物质本性的研究。20世纪90年代以来,一些信息工程师们开始着手研究如何将本体论的研究方法用于大型信息系统的开发,用于支持具有时间和空间跨度的大型组织任务的信息需求,从而发展出一套运用于信息系统设计领域的系统工程方法。在此,英文中的本体论(ontology)一词已经被引申为具有本体论研究特点的系统化方法,或直接指一种被设计出来的“本体系统”。所以,ontology一词在英文中使用,经常以“ontology engineering”或“formal ontology”的形式出现^[8,9]。

本体系统目前比较通用的定义为:“某一领域的概念化体系,及其对于概念之间关系的描述。”^[1]

这样一种系统概念,从文字上看来,似乎与一般的系统工程方法并无不同之处。但它的本质是人造的概念系统,核心的因素是人工规定的概念与规则体系。本体系统中的概念与一般的实体系统概念不同,它把刻画真实世界的实体概念、关系、属性等全部概念化了,并且加上一些与这些概念相关的领域公理,构成用于人工智能推理系统的逻辑基础。本体系统的应用条件,是相应领域专家对于本体系统概念与规则体系的共享,从而实

现专家群体与基于本体系统概念的人工智能体系的人-机衔接。

使用本体系统方法的目的,通常包括为系统提供数据库知识运用的能力,促进复杂系统和数据库中知识分享的能力,帮助人们更有效地理解和应用专门领域的术语,并帮助人们就专业领域的知识达成共识。

一般来说,本体系统要由对于特定专业领域具有充分理解的知识工程师来构建。在构成方法上,采用人工定义的概念,来表述系统的目标、属性及其内在的相互关系。这些人造的概念运用于一些陈述性语句、领域性公理之中,用来定义一个领域的整体。这些领域性公理,随后被用在推论体系中,成为知识和推论的来源,并成为决策支持的基础。

本体系统能够用于复杂非结构化问题的知识管理,是基于本体系统专家们的一些发现,即与真实世界对象的结构和行为相关的下述一般性研究结果^[2]:在某一领域存在一些对象;这些对象具有一些属性和意义;对象之间可以存在各种不同的关系;属性和关系可以随时间改变;事件发生于不同的时间瞬间;在一个时序过程中,存在对象参与的某一过程;各种对象及其所处的领域、状态可能多种多样;一些事件可以引起其他一些事件或者状态的发生;对象本身可能含有多个部件。

这些发现可以说是对非结构化问题的一般属性的描述。在此基础上,一些专门对本体系统方法进行研究的群体公布了构造本体系统的方法论框架^[10]:

- 1) 界定范围,决定本体系统包含哪些,不包含哪些因素。这要求对于领域进行研究,了解领域的边界和内在关系的始终。这意味着在将一个领域概念化和形式化之前,要对它进行充分的理解。

- 2) 适当地定义领域术语和名称,使得该领域的层级能够清楚地包含在名称的逻辑层级中。把概念指定成各种功能或某种判定,例如保证本体系统真实度的任意一种决策等。

- 3) 捕捉领域关于各种术语的通用知识,并用逻辑语句或领域公理的形式表述出来。这样可以帮助组织准确理解术语并藉此达成共识,并且可以为启动知识库的发展提供推理功能。

- 4) 通过写出与本体系统具体概念相关的简单

的公理语句,刻画具有专用意义的领域问题案例。

5) 提出假设来运行关于公理和具体问题事件的推理过程,并从本体系统中获取信息。用 Uschold 对本体系统构成进行的归纳性阐述来说明本体系统的基本构成:“一个本体系统可能采用不同的形式,但是不可或缺地,它应该包含一个有关术语的词汇表,和一些对于这些词汇意义的描述。其中包括定义和相互关系的意义的说明,它们共同构成了一个领域的基本结构,并对其中术语的可能性解释进行约束。”^[8]

3 基于本体系统方法的“电子公共大脑”设计

所谓“电子公共大脑”,是以计算机和网络为载体的,用以支持复杂群体自由交互的群体信息记忆和信息组织系统。该系统通过采用规范的多维信息定义方法及超文本连接技术,建立起逻辑运算体系,以实现非结构化自由交互过程的信息分类,并引导群体的注意力聚焦。

“电子公共大脑”的设计分为两部分:第一部分供群体共享对信息属性和关系进行描述的概念和约定,即本体系统;第二部分是在本体系统基础上建立逻辑运算、逻辑检索、信息组织、信息图形化系统。

3.1 电子公共大脑的本体系统

3.1.1 电子公共大脑本体系统的概念

引申本体系统的一般定义,电子公共大脑的本体系统,是人工约定的描述群体交互中信息属性的概念化体系,及概念之间的关系与规则的描述。群体信息交互过程中作为系统关注对象的所有个体信息、信息的属性及信息之间的关系,构成了本体系统所要刻画的对象。

3.1.2 电子公共大脑系统因素及其关系界定

电子公共大脑的系统因素界定,主要围绕信息表述中的一些特征变量和关系变量进行。在设计中提出下述供群体所共享的刻画信息属性和关系的概念^[4]:

信息名 每条信息应该有一个独一无二的信息名。这个信息名可以通过人工定义并由计算机系统确认其唯一性,或由计算机直接编码定义。

参与者 参与群体交互过程的相关人员,可以采用姓名、网址、团体名等,也可以匿名,同时可以根据需要进行相应的权重界定。

根据 信息交互中具体观点的相关依据或假设。作为一条信息的支持性文件,可以按照决策需要的规范决定其格式和文件名,在用于定义信息时,只采用它的地址或信息名,作为定义信息的要素,当然,这个信息名应该是独一无二的。

连接属性 反映信息与其连接对象之间关系的属性。

连接对象 观点表述的相关文本或连接对象。该要素采用连接对象的独一无二的信息名。一般作为群体研讨过程,必定要有一个研讨的主题,它的下面还可以有多个分主题,这些主题的信息名,就成为其他信息最初的连接对象。

时间 观点表述时间。

关键词 反映信息基本内容的关键词语。

权重 是反映信息重要度、参与者权威性的信息变量,它的规定与群体的组织机制、政治结构或领域划分有关,可作为群体信息交互过程引导群体注意力聚焦,或判别信息可信度的参考变量。

1) 关于信息交互中有关命题的相关根据(或假设)

在输入观点或命题时,如果不能提供支持其观点的相关依据或前提假设,就会产生理解上的知识断层;所以,对于正规决策过程中的观点表述,一般应该要求提供其相关的依据。

2) 关于观点表述的连接对象(或相关文本)

对参与群体决策的个体而言,任何一个参与人发表他的观点,按照决策过程的基本要求,这一观点应该与决策的问题相关。所以,任何与决策相关的观点在发表时,原则上已经存在一条与之相关的信息可以作为连接对象。这个连接对象可以是决策的主题,可以是相关的分主题,也可以是其他相关的观点或者支持信息等。

3) 反映信息与其连接对象之间关系的属性

“信息之间关系的属性”这个概念,就是信息之间关系的特点。信息之间关系的属性(或特点),决定于信息提供者、使用者本身的认识或者需求。这些属性,在不同的应用场合,可以有不同的定义方法和分类方法,本文仅讨论其中一种最普通且易于为人理解的分类型。

非结构化群体决策的信息交互过程中,个体信息活动的属性,即每个人在表述信息时能够反映信息之间关系的特点,包括态度、目的等,以此作为定义信息的一种属性分类,可显示出非结构化过程形成的大量信息之间的关系.这些属性可以包括表 1 中的变量^[5].

表 1 信息属性变量表

赞成	创新(或提议)	怀疑
反对(或异议)	中性(如:起支持作用的资料)	综合分析
质询	直觉	其它等等

这些属性变量,基本涵盖了通常个体信息表述可能出现的各种属性.这些属性,是反映信息之间关系的主要概念.

3.1.3 规则约定

信息交互的参与者在进行信息制作并输入群体交互集成系统时,必须定义“信息名、参与者、连接属性、连接对象、关键词”等反映信息属性的要素;“根据”这一要素可以依据该项信息是否还要用其他信息的支持来选用;“时间”由系统自动纪录并定义;“权重”由参与群体在本系统设计之外进行规定.

上述的概念和规则约定构成了公共大脑设计的本体系统.从上述概念及其内在关系的规定可以看出,这个本体系统的核心工作在于对系统要素及其关系、属性的概念化和对于人造概念及其描述的共享.

3.2 基于电子公共大脑本体系统的知识管理功能的实现

根据本体系统的概念体系和规则约定作为群体信息交互中信息制作的形式化表述模式,可以设计出十分丰富的信息管理逻辑功能.

3.2.1 信息结构化功能

首先,每一条信息根据它的形式化制作模式,它已经是与另外一条相关信息相互连接的一个单元环节.按照本体系统的定义,这个单元环节的基本结构见图 1.

按照图 1 的信息定义结构和连接方式,群体过程的信息输入(即便是头脑风暴法的非结构过程信息)随着时间的推移,都可以形成有序的树状结构(图 2).

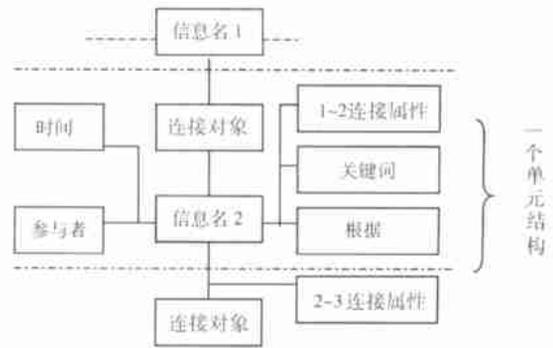


图 1 信息连接的基本单元结构

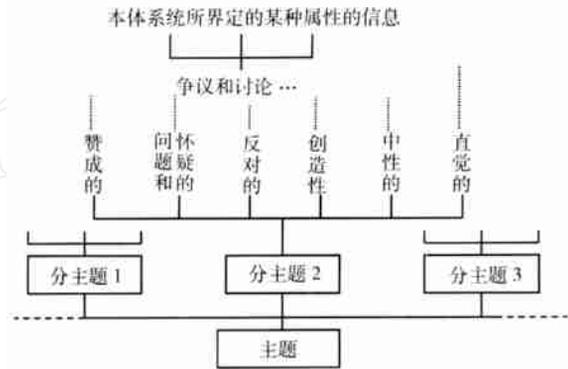


图 2 群体交互信息的树状结构

群体信息交互的过程中,当人们了解群体过程信息整体的、可以显示的树状结构时,虽然未必能够全面反映信息之间所有的关系,但是比起将群体过程视作无结构过程,这种树状结构可以提供帮助人们认识信息间关系的结构化途径.并且这是在自由的非结构化群体交互过程中可以即时形成的,最直接地描述信息整体和局部关系的结构化方法.

3.2.2 群体过程信息的定性集成功能

信息的多因素定义方法,加上信息输入时选择采用的权重定义(如:可以对于一些关键问题、权威专家、关键信息等赋予较高的权重),以及关键词查询等功能,通过一些逻辑运算,可以形成多种具有关键意义的群体决策定性集成和系统分析支持功能.

定性集成的主要任务包括从大量信息中提取比较重要的信息,并且通过这些信息意义、关系的分析与综合,得出能够指导下一步行动的方案.在网上进行的群体决策过程,大量信息中与信息重要性相关的判断因素,权重是最为普遍和容易采用的一种因素.权重与本文用于定义信息的多种

因素相结合,可以形成对非结构化群体过程十分有益的集成方法.下面通过各因素的组合分析,对其中部分定性集成功能及其意义进行列举.

假设一个群体过程对于一项决策问题中的分专题、参与人的重要度赋予权重,结合上述的信息定义结构,可以形成多种群体过程的定性集成功能:

1) 分专题 + 权重 对于各个分专题,根据权重予以检索分类,可以帮助参与人了解有关专题的重要度及其可能对决策产生的影响.

2) 问题 + 专题 + 权重 这里“问题”是连接的属性之一种,它与专题、权重的配合,可以帮助人们了解专题中由参与人提出的问题,及相关的重要性如何.利用权重作为逻辑判据进行问题排列,可以帮助人们重视关键的问题.

3) 争议 + 专题 + 权重 “争议”也是信息连接属性之一,它与专题、权重配合,可以引导人们注意群体过程中有哪些需要注意的冲突,相关的主题和重要度,并可以追溯相关的信息连接.

4) 问题权重分布 根据一定的权重,可以查询相应重要度的问题在信息树状结构中的分布.

5) 冲突权重分布 根据一定的权重,可以查询相应重要度的冲突在信息树状结构中的分布.

6) 参与人权重分布 根据一定的权重,可以查询相应重要度(或权威性)参与人的分布状况.通过与问题分布、冲突分布的结构,可以判断人员配置的合理性和有效性.

7) 专题 + 创造性 可以帮助查询各个专题中是否存在新的建议、创造性思路等.

8) 关键词 + 专题 可以帮助查询人查询相应的主题与哪些专题有关.

9) 关键词 + 参与人 可以帮助查询人了解相应主题与哪些参与人有关,有利于进一步交流.

10) 关键词 + 权重 可以帮助查询人了解相关主题在群体信息中的重要性如何.

对于电子公共大脑本体系统的信息定义结构,上面列出了 10 个明显的群体过程信息管理意义的信息检索途径.显然,这一本体系统中各种要素的组合还可以为更多的逻辑功能提供实现基础.它对于群体过程的支持作用可以归纳为下述几个关键的方面:

1) 作为中性信息制作工具 信息作为人工制品,采用本体系统语言进行制作,可以转换成不同

形式应用于复杂的目标系统.这样做的优点包括利于知识再利用,促进系统维护和利于长久的信息保存.

2) 作为对群体交互过程进行系统描述的工具

电子公共大脑本体系统可以作为对群体交互的系统过程进行描述和开发软件的基础.这一途径的优点包括拥有形式化文本信息,利于维护,可靠性高和利于知识再利用.

3) 作为公共信息访问对象 在很多情况下,信息是以不易理解或不易访问的语言及格式制作的,本体系统通过术语及其相互关系的知识共享,使得信息易于理解、查询和使用.

4) 作为群体交互过程的信息结构化和知识追索工具 群体交互过程中,每位参与者可能有不同的专长和感兴趣的领域,他们可以根据自己的需求,选择相应的专题,将自己的信息连接到群体信息的树状结构中去,并根据树状结构的信息连接,查询自己感兴趣的问题、冲突、创新性建议、支持信息、重要度、置信水平等内容.

5) 用于系统知识的维护和重复利用 电子公共大脑本体系统是对信息属性及其关系进行形式化编码的基础,可以在信息制作、软件系统设计中分享和重复利用.用清晰的本体系统方法构造的系统,可以服务于多种信息渠道、软件文本以降低维护的成本.特别当本体系统作为中性的编辑性语言服务于多种对象时,这种维护只要在一个地方进行就可以了.

3.2.3 电子公共大脑本体系统的系统特点与作用

电子公共大脑本体系统的构建原则,是根据信息的行为学意义,采用信息交互过程中与个体相关的行为属性要素构造概念系统,对信息进行定义,使信息成为超文本信息连接集合中的基本单元,并构成逻辑关系明晰、逻辑功能丰富的树状信息结构,为群体所共享.基于该系统的功能特点和工作过程由群体智慧所推动,可以形象地称之为“公共大脑”.群体共享信息不断增长和信息与信息之间逻辑关系不断建立和演化的过程,可以视作“公共大脑”的“思维”过程.当然在此“思维”过程中,人的参与是不可或缺的.其中的关键在于信息之间逻辑关系和内在行为关系的建立.

如此构建群体过程支持系统的优点,是不必过于深入地追究参与人思维过程的复杂性,而将

知识管理过程针对于参与人外在的、与信息表述相关的行为属性,帮助参与人在非结构化过程中形成群体决策的知识体系,并通过信息之间行为属性的逻辑连接,使决策者把握群体过程各种信息之间的联系.这个系统将群体决策过程中“人-人”之间的沟通过程,转变成“人-具有群体思维关系的信息(公共大脑)-人”之间的沟通过程.这

一过程可以帮助交互群体回避由于信息差异造成的面对面冲突,而将信息的共享与沟通作为主题,将个体信息纳入群体共有的信息结构中,供群体过程的参与者进行演化.不同参与人之间存在的观点、问题和矛盾可以构成信息沟通和信息演化过程的基本动力,为群体决策过程提高决策质量提供源头活水.

参考文献:

- [1] Gruber T R. A translation approach to portable ontologies[J]. Knowledge Acquisition, 1993, 5(2):199-220
- [2] Chandrasekaran B, Josephson J, Benjamins V. What are ontologies, and why do we need them? [J]. IEEE Intelligent Systems, 1999, 14(1):20-26
- [3] Uschold M. Knowledge level modelling: Concepts and terminology[J]. Knowledge Engineering Review, 1998, 13(1)
- [4] 程少川,张朋柱,王刊良.电子公共大脑的信息组织设计研究[J].西安交通大学学报(社会科学版),2001,(1):19-24
- [5] 程少川,孙景乐,卢明德.群体过程信息的树状结构及其定性收敛的研究[J].系统工程学报,2001,16(5):371-375
- [6] 程少川,王刊良.基于行为的思维分析和DSS中信息组织方式的研究[J].系统工程理论与实践,1995,15(11):31-37
- [7] 王刊良,程少川,席酉民. GDSS环境下群体沟通的研究[J].决策与决策支持系统,1995,5(2):21-26
- [8] Uschold M, Healy M, Williamson K, et al. Ontology reuse and application[A]. In Guarino N ed. Formal Ontology in Information Systems[C]. Trento:1988.179-192
- [9] Robert Jasper, Mike Uschold. A Framework for Understanding and Classifying Ontology Application[R]. Proceedings of the IJCAI-99 Ontology Workshop,1999
- [10] Gruber T R. Toward principles for the design of ontologies used for knowledge sharing[A]. Nicola Guarino. ed. Formal Ontology [C]. Padua Workshop,1993

Study on artificial intelligence of knowledge management based on ontology

CHENG Shao-chuan¹, CHENG Xiang-qian², ZHANG Peng-zhu¹

1. Management School, Xi'an Jiaotong Univ., Xi'an 710049, China;

2. Center of Computer Teaching and Experiment, Xi'an Jiaotong Univ., Xi'an 710049, China

Abstract: This article analyzed the dilemma of information organization for knowledge management, and the essential of unstructured attributes of group process. It introduced the concept, construction, and application of ontology for unstructured problems in artificial intelligence, through a typical example of system design that is called Electronic Common Brain, which is used for knowledge management of group communication.

Key words: ontology; electronic common brain; knowledge management