

# 知识表达与超文本

## Knowledge Representation and Hypertext

罗春晔

Luo Chunye

(广西计算中心广西软件新技术实验室 南宁市星湖路 32 号 530022)

(Guangxi New Software Technology Lab., Guangxi Computing Center,

32 Xinghu Road, Nanning, Guangxi, 530022)

**摘要** 知识代表对象间被认识的关系, 这些关系使自动推理或至少是有效的逻辑运算成为可能。有效地利用知识可以实现超文本的智能, 超文本的知识表达是实现智能的超文本的关键。

**关键词** 超文本 知识表达 人工智能 专家系统

**Abstract** The word knowledge stands for elicited relationships between objects making possible automatic reasoning or, at least, efficient logical computation. Knowledge representation is the key technology of an intelligent hypertext which could be realized by exploiting knowledge.

**Key words** hypertext, knowledge representation, artificial intelligence, expert system

### 1 前言

知识代表对象间被认识的关系, 这些关系使自动推理或至少是有效的逻辑运算成为可能。而信息更多是面向用户的, 它代表人们可能感兴趣的任何事物<sup>[1]</sup>。可以说, 知识是被认识了、条理化、信息, 知识好比树干, 信息好比树叶。

超文本离不开知识, 因为超文本技术本身就是用于信息管理和知识管理的, 少不了要和信息、知识打交道, 而且, 利用与人工智能有关的知识可以实现超文本的智能化, 解决超文本在导航和检索中遇到的困难。

如何有效地表达知识是实现智能的超文本的关键。

### 2 超文本的知识表达<sup>[2,3]</sup>

超文本实际上是一种在许多方面与语义网相类似的极为灵活的知识表达环境。象语义网一样, 超文本也是由节点(对象)和链组成。不同类型的知识表达形式都可以在超文本中通过以不同的方法对节点和链的基本类型进行结构描述和定义来实现。以下介绍超文本中最常见的节点类型和链类型。

#### 2.1 节点类型

超文本的节点可分为表现类节点、组织类节点以及智能超文本中的推理类节点。

表现类节点包括文本节点、图片节点、声音节点、动态图象节点、混合媒介节点和按钮节点。

• 文本节点 它由文本块组成。文本本身可以是一份文档或被定义成一个由节点表现的对象; 也可以表达基本信息。例如一份文档中所提供的文本可以表达未被抽象成知识表达形式的信息, 读者可以从中提取知识。

• 图片节点 图片节点可以嵌入文本节点中也可以是独立的节点。图片和文本可以互相补充说明: 可以用一幅画去解释达芬奇传, 或者在看完一幅图片后, 再看看艺术家传记中的详细描述。带放大链的图片可以用于更细致的检查。象文本一样, 图片表达非解释性的信息, 而不是知识。

• 声音节点 象图片一样, 声音可以嵌于文本中, 或做为一个独立的节点而存在, 它与图片表达方式相似。声音也表达非解释性信息。

• 混合媒介节点 这种节点是文本、图片、声音的某种组合。在许多情况下, 同一信息可以用被链接的节点组合表达, 或是用单个混合媒介节点表达。象屏幕大小、超媒体系统中沿链浏览的难易程度、审美考虑、内存的分配以及用户的专注能力等因素都将影

响决定是否将信息分成多个节点。

• **按钮节点** 一条链连接一对节点,而一个按钮执行一个进程。带按钮的节点常被用来做某件事情。这些节点允许超媒体做为一个高层的程序或数据库界面。复杂的超媒体将普遍采用带按钮的混合媒体节点以提供最大的灵活性,上述4种信息节点是这种更普遍节点类型的严格版本。

以上介绍的5种节点用来表达信息,而非解释性的知识。

**组织类节点**是组织节点的节点,加索引是描述节点的一种方法,同时也是数据库管理的需要。组织类节点可以实现数据库的部分查询工作。组织类节点包括各媒体的目录节点和索引节点。

• **目录节点** 目录节点包含各自媒体的索引指针,指向索引节点。

• **索引节点** 这些节点由索引项组成,索引项用指针指向目标节点,或指向索引项的相关索引项,或指向相关表中相对应的一行,或指向原媒体的目录节点。

**推理类节点**用于辅助链的推理与计算,它包括对象节点和规则节点。

• **对象节点** 这种节点描述对象,它们由槽、继承链、相关进程组成。对象节点和 is-a 链能够用来表达组织性知识。

• **规则节点** 这些节点列举规则,并且可以指向满足规则的对象、采纳规则的理由、规则的使用说明等的组合。

## 2.2 链类型

可以在超文本中定义不同类型的节点,使之成为一个极灵活的知识表达工具,这种灵活性将通过定义不同的链类型得到更大的加强。链定义了超文本的结构,并提供了浏览、查询节点的能力。超文本有许多不同类型的链,它们可以划分为导航链、推理链及组织链。

### 2.2.1 导航链

在这儿我们来看看导航链的四种主要类型。这些链的作用可以通过类似于摄像机的操作来理解。作为导航实体的链对应于在一个显示屏上景观的改变,这与摄像机的操作改变可见景观的外观大致相同。人们可以先简单地前后移动摄像机(移动链),然后调焦在景致的一个特殊部分(放大链),然后又摇动镜头回到大图(缩小链),最后,人们可以用不同的过滤镜去高亮显示景物中的不同方面。在超文本中可以用视图链使链决定于上下文限制的特殊集合中。这种导航链的4种类型如下:

• **移动链** 这是一种最基本的导航链。沿这些链浏览可以访问到一个相关节点,它们允许围绕超文本移动或在超文本中导航。

• **放大链** 这些链把当前节点扩展到一个更详尽的信息节点中,其效果类似于从一个摘要移到本文档的全文,或是类似调焦得到一个全图的一小部分的放大图,更类似于从一个全市地图移到与全市地图等大小的市中心地图。

• **缩小链** 这种链返回超文本的一个更高层的视图(这些链在浏览中特别有用)。缩小链正常情况下与放大链相反,所以放大链将对应一个缩小链,反之亦然。这样,这种链也可看作将镜头摇远。

• **视图链** 这些链的有效性与其灵活性决定于用户要说明的目的和用户的兴趣。它们也可用于保险目的。视图链是隐藏的,除非某个用户对它们很感兴趣。例如,在水星和行星节点之间有一条视图链,它对天文学家是显露的。但对化学家和气象学家却是隐含的。视图链可提供基本的机制用于按不同用户的需要与兴趣裁剪超文本。它们也帮助防止超文本不必要的复杂性,使用户不被那些与他们需求无关的信息所纠缠。

### 2.2.2 组织链与推理链

组织链和推理链用于节点的组织以及和更通用机器推理、编程能力的链接。这里侧重于介绍五种组织、推理链类型。

• **索引链** 这些链将用户从一个索引节点引到该节点的相应索引项。索引然后被用来进入关系数据库或查找共享特定索引项的文档。索引超文本是控制节点间的链的激增的好办法。

• **Is-a 链** 这些链类似于用在语义网和对象系统中指示一个类型的成员的链。

• **Has-a 链** 这些链用于描述节点特性,它们通常被对象使用。

• **隐含链** 这些链用来连接推理树中的事实,它们一般等同于正在使用的规则,或过时的规则,隐含链的使用通常限制在推理超文本中。

• **执行链(按钮)** 执行链将一种操作与按钮节点相连,使超媒体成为一个更高层程序的界面。

## 3 利用知识实现超文本的智能<sup>[4]</sup>

当超文本系统变得庞大以后,导航机制和检索机制的薄弱就暴露出来了,如何解决这个问题呢?关键就是引入人工智能技术,开发和利用有效的知识。

要在一个大文档中查找读者所需的某些信息似乎是一件毫无希望的事情,但人们能够而且也确实做

到了。关键技巧就是读者自然而然地应用了各种各样的知识去帮助缩小查询范围。以下是其中 3 种重要的知识形式：

- 组织性知识：说明文档的内容如何组织。
- 专业知识：包含在文档内容中的明确的知识。
- 上下文知识：说明使用文档信息的方式。

以下我们介绍如何把以上讨论的几种知识应用于超文本中。

### 3.1 嵌入组织性知识

组织性知识描述文档的组织结构，它与文档内容无关，所以关于面包制作的文档可以和神经外科的文档有着同样的组织结构，一些组织性知识中的典型的描述如下：

这一节需要阅读。

阅读这一节以前首先需要完全理解第七节的内容。

第一次阅读时这一节可以跳过去。

如果你跳过了第三节，那么就跳过接着的五个章节。

如果读者没有选择“示例”，就从文档当前显示中略去指导信息。

本节是引导章节。

传统文档中，这种结构的语义描述会在文中明确表示出来。比如有些数学论文，难懂的章节用“S”作标记，因为“S”在路标中意味着“前方道路曲折”，每一章的前言列出预读章节的清单。

在一个超文本文档中嵌入这类含语义的结构，就能从显示中发现这些可以利用的关系。比如，软件知道你还没读第三节，它就会隐藏紧跟的五节（直至读了第三节）。

现在的关键点就是把结构的语义用“如果——那么”的规则表示出来，简单的说明（“第七章很难”）可以用这个普通的规则代替：

如果 真  
那么 “第七章很难”

因为真永远是真，所以第七章永远很难。文档的结构语义可以用更普遍的应用规则来开发：

如果 读者一直在阅读指导手册，而且他也  
一直没读参考手册

那么 该读者显然是一个初学者

如果 该读者是一个初学者

那么 跳过较难的章节

很显然，好的文档作者总是试图把难度大的资料集中在某些章节（加上警告注解），以此来协调推理的进行，但是，正如我们所看到的，当这种方法应用

到纸张文档时，它就比较被动，因为读者必须按作者的提示阅读。

使用计算机文档的一个麻烦就是要花时间去查找一些实际上你不想知道的信息。如果文档知道你最可能不想读它，那么它就不应该让你去查找或者表明它是可读的。

结构语义中也嵌入了路标的观点，一个路标就是一个特别重要的章节。路标是一个醒目的标志——当读者在某点迷路，就可以把他带到最相关的路标处，路标使读者在纷乱复杂的环境中获得一种稳定感。

### 3.2 嵌入专业知识

组织性知识涉及文档的组织结构，它与文档内容无关，而专业知识则涉及文档的内容。专业知识告诉你文档说什么，而组织性知识告诉你它是如何说的。

在计算机系统里，处理专业知识比处理组织性知识更难，因为它主要依靠人们认为理所当然的常识。让计算机系统演示常识是当前的热门课题，彻底解决它还需要很多年。

我们可以利用推理式的规则来描述专业知识，例如：

如果 研究课题是巴洛克音乐  
那么 把研究限制在 1550 年至 1750 年这段  
时期

表达这条规则的另一种方法就是在西方文化中定义巴洛克音乐的时期。

这种方法的问题就是专业知识实在太丰富了，结构的语义能缩减为一个管理文档的结构和指定文档的特定事实的规则集合，但专业知识则无边无际，所以我们需要表达能力更强的工具。

一个相关的 AI 技术就是概念树叉，它是一个能表示某些专业知识的主要概念和概念之间关系的数据结构。概念树叉能用这样的要求帮助用户理解文档的内容：

给我一个主要问题的更详细的讨论  
把这一节放到一个更普通的上下文去

我们不必为每一个要出版的超媒体文档发明一种新的概念树叉，重复使用一个概念树叉就足够了。另一方面，如果概念树叉都要覆盖大量资料，那就要冒缺乏实际内容的危险。

Douglas Lenat 和他的同事们正在 MCC 国际财团做着相关的工作，他们正在建造一个智能百科全书系统 Cyc。Cyc 将由大量的生活中的事实（“玻璃是透明的”，“公开裸体让人恶心”）和常识（“玻璃屋中的人必须穿衣服”）组成。建立 Cyc 系统是一个巨大的任务，而且利用它建立的应用系统也需要一段很

长时间。但毫无疑问,它肯定能用更一般的结构取代书写的专业知识语义描述。

概念树叉也为文档提供了有效的图示,它说明了文档的内容是如何组织的,而不象大多数超文本关系图说明文档是如何组织的。这两者有很大的不同。在特殊情况下,如果你正处于超文档的边缘(节点之间没有链相连,且距离很远),你就会走到一个作者不能带你出去的死胡同。为了找到边缘信息的相关信息块,可以利用关键词检索找到相关信息,然后再沿着链浏览,找到想要的块。这两个块在概念上是相关的,所以在概念树叉中距离很近,但你从浏览的链图中无法推测出这种关系。

### 3.3 地方代理

结构或专门知识的语义把用于帮助读者检索信息的知识嵌于文档之中,你需要这些知识,不管是用于结构语义的规则的形式,还是以用于专业知识语义的概念树叉的形式组成,同样重要的是你要有一定的开发知识的能力。

解决这个关键就是代理的概念。代理是这样一些程序,为读者利益服务,能实现读者的计划安排,可以使用嵌于文档的知识和返回的信息。可以把地方代理看作是文档作者的代表,能回答你一个读者想知道的问题:

我该先看哪一章呢?

能给我显示最重要的章节吗?

地方代理包含应用文档中的规则进行推理的推理机,它必须提供一个良好的用户界面(可能是图形的或是自然语言驱动的),接受用户请求,做出反应,它也负责协调总代理的工作。

### 3.4 总代理

当一个固定的文件集被庞大的动态的文件网所代替,超媒体文档的作用将被充分认识。网络的概念正顺应了当前计算机系统向容量共享、高速网络连接、高度分布化发展的趋势。

在现有超文档中增加一份文档可能是有问题的,新的文档必须适合用概念树叉表示,而且它们的链必须建立在过去出版的文档基础上。

地方代理是为超文档的读者服务的,总代理是为动态网服务的。当增加文档时,总代理就在概念树叉的驱动下链接信息,然后通知地方代理网络的最新状态。今天的分布式数据库系统当增加或删除网络的数

据库节点时能准确提供这种功能。

因为总代理要管理整个网络,所以它们应能看到网络的使用情况,这是在计算机上建立文档时所利用的另一种重要知识。另外我们能用也应该使用一些明确的、独立于专业知识的规则。比如:

- 沿着走过的路径浏览。

如果一个读者询问:“下一步我们去哪儿?”而且你又没有一个好的答案,就可以带他到大多数人去的地方。

- 地球在转动,相互关系也随着时间和年龄变化而变化。

如果读者不满意以概念树叉方式使用网络,那么树叉就过时了,必须改变它。

总代理密切注视人们使用网络的情况,当总代理注意到某个模式重新发生时,他们能帮助地方代理和读者发现有关信息。地方代理可以暂时成为其他代理的用户,当一个读者提出一个地方代理不能解决的信息需求,代理会向总代理寻求帮助,总代理知道去哪儿解决问题,也能调用一个或更多的地方代理为它提供所需信息。

## 4 小结

从前面的讨论可以看出知识在超文本中具有举足轻重的地位,超文本不仅仅是信息浏览的工具,它更是知识管理的工具,而且知识的有效利用可以使超文本具有一定的智能,解决庞大的超文本系统的导航、检索问题。超文本和其它信息管理技术相比,有更灵活的使用方式,有更准确、丰富的知识表达,也就有更广泛的应用领域和应用前景。

### 参考文献

- 1 Nanard J, Nanard M. Using Structured Types to incorporate Knowledge in Hypertext. Hypertext '91 Proceedings, 1991.
- 2 Parsaye K, Chignell M, Khoshafianand S, Wong H. Intelligent Databases. John Wiley & Sons, Inc., 1989, 242~258.
- 3 老松杨, 胡晓峰, 张茂军. Hypertext 系统的结构与特征. 计算机世界周刊, 1993.
- 4 Littleford Alan. Artificial Intelligence and Hypermedia. Hypertext/Hypermedia Handbook. McGraw-Hill, 1991.