

面向专家的知识获取

曹存根 著



科学出版社

991571

面向专家的知识获取

曹存根 著

科学出版社

1998

内 容 简 介

本书介绍一种面向专家的知识获取方法。这一方法的出发点是：知识获取的中心并非知识工程师本人，而应该是具有丰富实践经验和较高认知水平的领域专家。根据这一论点，不难发现，知识获取困难的本质在于三个相互关联的切入问题：即心理切入、认知切入以及知识切入。一旦知识工程师处理好这三个问题，知识获取也就进入了正确的轨道。本书共分九章，将着重分析知识获取的初期阶段，探讨知识工程师如何尽快地、恰当地切入到专家领域以及专家认知过程中的方法、途径。

本书适合于从事知识获取实践和研究的专业人员阅读。由于知识工程师将成为一个越来越重要的社会职业，因此在写作时笔者尽力使该书成为一本知识获取的入门书，以供那些具有一定计算机知识并且打算从事知识获取工作的大专生和本科生阅读。

图书在版编目(CIP)数据

面向专家的知识获取/曹存根 著. —北京：科学出版社，1997

ISBN 7-03-006145-4

I. 面… I. 曹… 1P18

中国版本图书馆CIP数据核字(97)第14000号

科学出版社出版

北京东黄城根北街16号

邮政编码：100717

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经营

*

1998年1月第 一 版 开本：787×1092 1/32

1998年1月第一次印刷 印张：6 1/8

印数：1—1700 字数：125 000

定 价：13.00 元

献给

云南师范大学数学系刘声烈教授

EAB 14/01

知之为知之，不知为不知，是知也。

—孔子

知识和错误是同一硬币不可分割的两面，只有成功才能区分它们。

—Ernst Mach

序

知识工程作为一个研究领域约在 80 年代初传入我国。当时开发的关幼波肝炎诊治专家系统也许是国内第一个实际投入使用的专家系统。自那时以来，知识工程研究在我国有了长足的进展。随着一大批专家系统的出现，这个领域在我国逐渐走向普及。然而，曾经有一种看法认为专家系统不过是计算机软件的一个普通的新应用领域，并不需要用专门的技术和工具去对付它。我曾对 80 年代中期的一次国内医学信息处理会议作过统计。在该会上发表的大量专家系统文章中，有百分之六十以上的专家系统是用 BASIC 或汇编语言这类低级语言编写的。少数专家系统用 PASCAL，FORTRAN 或 ALGOL 等一般高级语言编写。只有个别系统用了人工智能语言 LISP。用专门的知识工程工具开发的则完全没有。这种情况反映了当时部分学者对知识工程这门学科的认识，也反映了当时国内的技术水平。

80 年代中期以后，有关专家系统开发技术和开发工具的研究正在兴起。除了引进国外的专家系统开发工具外，国内也出现了一批自行开发的工具。尽管其中商品化的不多，但有信息表明我国现有的专家系统基本上都是用国产工具开发的。同时，有关专家系统和知识工程的论文和著作也大量出版，使人们听到了我国在这个领域追赶国际水平的步伐。

在仔细研究了现有的专家系统开发工具（其中也包括国外的）以后，我们发现了一个比较普遍存在的问题：它们的

功能大多集中在知识库构造, 推理编程, 和优美的人机界面上. 对于知识工程的公认瓶颈——知识获取, 则下功夫甚少. 实际上, 根据许多专家的统计, 知识获取的工作量常常占整个专家系统开发工作量的百分之九十左右. 其结果是, 工具的使用并不能使开发者避免大量繁琐的知识获取工作. 有些所谓很快开发出专家系统的报道若不是因为那个专家系统很小, 就是因为没有把知识获取的工作量算进去.

我们认为, 知识获取是一个非常值得研究的课题. 我们自己多年来的研究实践也说明了这一点. 有些专家系统开发工具的推销商把他们的工具带有的知识编辑界面说成就是知识获取. 这是片面的, 甚至是一种误导. 知识获取的范围非常广泛. 正像知识工程有它的生命周期一样, 知识获取也有它自己的生命周期. 从早期的直接从原始资料(人类专家和书面材料)中获取知识, 到中期的从知识素材中提炼规律性知识, 到后期的通过实践检验和修正知识, 经历着无数次周而复始的循环. 专家系统和人一样, 也应该“活到老, 学到老”. 从这个观点来看, 目前国内专论知识获取的著作极少(或没有), 是和知识获取研究的重要性不相称的. 国际上不仅有了这个领域的许多专著, 而且有了专门的刊物和国际系列会议. 我们需要赶上去.

曹存根博士曾在我的课题组从事知识获取研究多年, 做出了很优秀的成果. 他把他的研究成果总结成书, 我很赞成. 他在本书中强调的观点, 即应以领域专家为中心来获取知识, 我认为也是很有道理的. 我个人并不担心这本书的读者少, 相信广大的知识工程、人工智能和认知科学研究者会对它感兴趣. 我还期待着它对推动我国的知识获取研究会起到良好的作用.

陆汝铃

1997年3月于北京

前 言

1. 知识获取的意义与困难——众家之说

纵观人工智能近 40 年来的发展历史，我们不难发觉，人们在看待人工智能时的心情十分复杂。在 1995 年人工智能被正式确定为一门学科的前后几年，人们研制了一些具有一定智力水平的跳棋、象棋以及其他游戏程序。许多人认为这些程序是未来会思维的机器的原型。特别是，美国人工智能大师 Simon 和 Newell 在 1958 年还提出了至今仍被一些人当作笑柄的三个预言：用不了 10 年，计算机将成为世界象棋冠军，除非另设竞赛规则，不允许计算机参赛；用不了 10 年，计算机将发现并且证明到那时还未被证明的重要数学定理；用不了 10 年，大部分心理学理论将在计算机上形成。

人工智能取得的一些成就以及遭到的一些失败，使得人们开始以种种形式关注它的命运，各种关于人工智能的争论也相应地出现。1971 年，IEEE 下属的一个协会组织了力量，分析人工智能的作用以及局限性。最终的分析报告指出，有相当一部分人认为人工智能最终可以实现。持这种观点的人有 Simon, Newell 和 Minsky 等。他们的理由大致是：所有的人类思维活动都可利用信息处理程序的语言在形式的神经网络平面上加以描绘，而有关思维的动机和情感等均被视为与思维无直接关系，甚至被看成使问题求解过程复杂化的因素。但是，也有相当一部分人持相反的看法。他们认为，人工智能的热衷者们显然是对人的智能理解得过分天真。他们

过分夸大了信息技术的作用，对人的心智活动缺乏深刻的认识。更进一步地，他们还认为人工智能与人的思维没有本质上的相同之处。

在人们关于人工智能进行各种争论的时候，美国斯坦福大学的人工智能学者们在 60 年代开展一项研究计划，即 Dendral 工程。他们研制出的 Dendral 系统包含了大量的化学分子分裂规则，根据这些规则以及质谱仪和其他实验数据，Dendral 系统推导出未知化合物的分子结构。在实际的使用中，Dendral 取得了令人瞩目的成就。

与此同时，Dendral 的研究者们也在深刻地思考着知识在人工智能系统中的意义。他们的结论是，人们无法找到或设计出通用而又有效的问题求解方法；有效的问题求解有赖于问题领域的专门知识。这一观点与传统的人工智能观点有着本质的不同。正如 Feigenbaum 1984 年再次强调的那样：长期以来，人工智能研究几乎无一例外地集中在探讨聪明的推理方法上，几乎所有的方法都能解决问题；但是，人工智能的力量来源于知识。不难看出，人们开始重新认识英国近代杰出的经验主义哲学家 Bacon 的一句名言：知识就是力量！

随着人们认识论上的改变以及一些相应成果的问世，一门新的分支从人工智能中诞生了。它就是我们现在所熟悉的知识工程，由斯坦福大学人工智能专家 Feigenbaum 于 1977 年在美国召开的第五届国际人工智能联合会上正式提出。

在知识工程诞生的前后不久，人们取得了许多令人兴奋的成就，研制成功了一些著名的基于知识的系统（如 Dendral, Mycin 以及 Prospector 等）。这些成就使得一些乐观主义者尝到了“知识就是力量”这一名言的甜头。但是，许多知

识工程学者除了尝到甜头外，还同时体味出了这一名言背后的苦涩，因为他们尚不知道如何有效地获取问题求解所需的领域知识。

Bacon 在他的名著《新工具论》中系统地讨论了归纳法与知识的形成等问题。他认为，只有通过归纳并依靠经验，人们才能从对外部环境的观察中获得一般性的知识。我们且不说这一观点的正确性如何，试问归纳所依赖的经验从何而来呢？如果按照 Bacon 的说法，经验还是来源于对观察的归纳，那么很显然我们陷入了一个无穷无尽的归纳过程。而这一过程的最初状态又是什么呢？我们无法满意地回答这一问题。因此，Bacon 虽然指出了“知识就是力量”这真理，但他的归纳法不能成为令人满意的知识获取方法。只有当人或机器具有了足够多的经验后，归纳法才能发挥它的作用。

为了加速知识获取的进程，人们已经从各个角度研制了许多知识获取方法和工具。知识获取的研究和实践已有 30 年的历史。回顾这一段历史，我们不难发觉，知识工程研究人员在研究知识获取时的心情也很复杂，因为他们碰到了许多棘手的问题。从事过知识获取实际工作的人也许都有一个共同的认识：对于知识获取的困难，无论怎么强调都不会过分！

读者不禁要问，知识获取究竟难在哪里？困难的实质又是什么呢？目前，人们对这些问题已经进行了深入的探索，但是所得到的答案不尽相同。

Hayes-Roth、Waterman 和 Lenat 认为，知识工程师的职责在于在专家和专家系统之间扮演中间人的作用。但是，因为知识工程师拥有的领域知识远远少于领域专家，所以知识工程师和专家之间的交流就成了知识获取的障碍。而在与

外行交谈之初,专家所使用的术语常常是不精确的,因此知识工程师必须和专家一道来精化和扩充那些不精确的术语.知识工程师工作中最困难方面之一是帮助领域专家对他的领域知识进行结构化,识别领域概念并对它们进行形式化.

Roth 和 Woods 认为,“自己依靠自己的力量”是知识获取的本质困难,也即知识工程师必须预先具备大量的领域知识才能向领域专家提出恰当的问题以及观察和理解专家的行为,而知识工程师往往不具备大量的领域知识.要具备大量的领域知识,知识工程师必然花费很多的时间阅读一些与专家领域有关的书籍资料,了解领域的概况以及一些术语、概念和过程的含义.在了解专家领域的过程中,知识工程师还免不了向专家求教.

Lenat 认为,知识获取困难的关键在于知识转移(knowledge transfer).它包括两层意思:一是专家难以表达他们的知识,二是知识工程师与领域专家之间没有共同理解的概念和表达这些概念的共同语言.

Gruber 指出,专家知识表示形式与知识系统所需知识的表示形式不匹配(representation mismatch).如何处理这种知识表示上的不匹配就成了知识获取的中心问题.

Johson 从心理学的角度有力地指出,随着专家为有效和准确地解决问题而努力掌握一门或多门专业知识的同时,他们也在不断地丧失他们究竟拥有哪些知识的意识.因此,知识工程师所要获取的知识可能正是专家难以表达的那些知识.

Steels 认为,如何向专家提问是知识工程中最困难的问题,提问经常被视为瓶颈.现在已清楚,专家很难以规则的形式给出他们的知识.即使他们给出了规则,它们通常也是

不完整的、太过于一般化或者太过于特殊化。而且，专家常常难以提供或者常常忽略掉一些重要的规则。在许多情况下，专家们之间对一些规则的看法也是不同的。专家口语记录中常常省略一些推理步骤，而知识工程师要填补省略的步骤是很困难的。

Lu 列出了知识获取难点的一个清单，它反映了知识获取各个阶段出现的难点。这一清单足以使人相信知识获取确实不是一件轻而易举的事。Lu 认为，知识获取至少包括以下几方面的困难：

- (1) 知识工程师要学习一门几乎全新的领域知识；
- (2) 知识工程师要寻找一位或多位具有丰富知识的领域专家，但是许多专家将他们的某些知识视为秘密而不愿公诸于众；
- (3) 知识工程师要尽可能多地从专家脑子里挖出知识，但是不是每一位专家都是一名好教师；
- (4) 专家所提供的知识往往是较粗的，而且经常含有个人偏见甚至错误。知识工程师必须对这些粗材料进行筛选、测试和重新组织；
- (5) 知识与经验不同，经验并不总是可表示的，它们是模糊的和不一致的。专家往往也很难解释经验，知识工程师更难理解它们；
- (6) 领域知识与常识之间没有清晰的界限，常识是非形式的、无限的、连续的以及无处不在的。知识工程师很难判定应该获取哪些常识；
- (7) 知识获取不是一蹴而就的，它需要知识工程师在很长一段时间内才能完成。

且不说上面所提及的有关知识获取困难因素是否完全和

系统，它们已经使许多热衷于知识获取研究和实践的人常常产生悲观情绪。但许多知识获取研究人员知难而进，研究了大量的试验性知识获取方法和工具。

2. 谁是知识获取的中心：知识工程师？领域专家？

这一标题不禁使我们想起了 16 世纪的哥白尼(Copernicus)。他曾尖锐地指出，宇宙的中心是太阳而不是地球；地球及其他行星都围绕太阳作圆周运动。尽管这一观点在今天看来不完全正确，但是它否定了人们头脑中根深蒂固的错误信念——地球是宇宙的中心。

我们发现，长期以来，在知识获取研究和实践中也存在着类似的问题，即知识工程师应该以领域专家为中心来开展知识获取，还是领域专家跟着知识工程师“转”？

在详细分析了许多著名知识获取方法后，我们认为，现在的知识工程师以及他们所使用的知识获取工具并不是真正面向专家的，知识工程师常常有意或无意地以自我为中心，他们不愿意或者不懂得如何站在领域专家的角度来思考知识获取问题。因此，知识工程师与领域专家之间原有的鸿沟又被知识工程师的这种态度变得更宽。那么，什么样的知识工程师才是面向专家的知识工程师呢？换句话说，知识工程师怎样才能做到站在领域专家的角度来进行知识获取呢？

我们认为，关键问题就在于三个层次不同但又相互关联的“切入”，即心理切入、认知切入和知识切入。心理切入是指知识工程师应该深入了解领域专家对知识工程师本人、专家系统技术以及知识工程师所使用的知识获取方法和策略等是否能够理解和接受。当然，我们这儿所指的理解是一种抽象层次上的理解，并非设计或实现层次上的理解。认知切入

从另一个角度认识专家本人及其领域. 它指的是知识工程师应该从认知的层次上了解专家对领域问题和知识的认识. 知识切入是指知识工程师切入到专家领域的问题求解知识中去. 正如读者将要看到的那样, 这三个切入问题是一个有机的整体, 我们不能人为地将它们分割开来.

在知识获取的整个过程中, 上述三个切入问题呈现出不同的形态, 并且表现出不同的难度. 经验表明, 在知识获取的初期, 这三个切入问题尤为关键. 它们的关键性主要体现在两个方面. 第一, 如果对他们处理得不恰当, 那么知识获取过程可能会夭折; 这一点在知识获取实践中是常常碰到的. 第二, 退一步说, 即使知识获取过程不会夭折, 知识获取前期的问题和困难也会传播到后期去, 从而人为地加大后期知识获取任务(如知识求解、知识库测试以及知识库验收)的难度.

上面所表达的观点与常识“万事开头难”同出一辙. 本书的重点就是分析在知识获取初期, 心理切入、认知切入和知识切入这三个问题的具体表现形态(即这三类问题所直接或间接表现出来的具体问题), 讨论它们的解决策略和方法.

在知识获取初期, 领域专家通常会考虑哪些具体的方面呢? 尽管这一问题的答案与具体的领域以及与具体的领域专家有关, 但是我们从专家的角度总结了三类重要的方面, 它们通常使领域专家感到疑惑不解.

第一是专家系统方面的问题. 对专家系统缺乏了解的领域专家关心的问题有: 何为专家系统? 它真的像人一样会思维吗? 谁是系统用户? 等等. 当他们在知识获取过程中逐步发现专家系统并非圣贤时, 他们会顾虑到另一个更严重的问题: 专家系统出错会给自己的权威造成什么负面的影响?

第二，领域专家多年的领域实践和专业修养才使得他们成为今天的专家，他们会担心与知识工程师这种特殊的外行能否进行有效的合作。由于他们的自由时间有限，他们也不得不考虑如何抽出时间与知识工程师进行合作。他们的正常工作在一定程度上会受到影响。他们也不免关心知识获取与他们本职工作之间有什么关系。

第三，正如前面所言，领域专家的长期实践与修养已使他们在很多方面区别于一般的领域人员。他们善于从整体上把握问题，而不是片面地认识问题。因此他们也期望知识工程师能从整体上把握领域知识。但是，他们又会怀疑知识工程师能否做到这一点，因为知识工程师对他们的领域知之甚少，或者仅仅略知一二。知识工程师进行知识获取的一个重要目的是获取专家的个人经验，他们也时不时地询问专家哪些内容是他们的个人“绝活”或“看家本领”。但是，正如我们多次经历过的那样，领域专家常常说不出来个人经验与公共知识之间的界限。因此，知识工程师所获取到的知识可能并非专家的真正知识。专家关心的另一个问题是知识工程师如何处理一般性知识与具体性知识。同时在专家看来，一般性知识的应用与具体的问题环境有关，不应该照搬一般性的知识来解决具体的问题。由于知识工程师缺乏具体的领域实践，领域专家会担心知识工程师能否把握住具体知识与一般知识的关系。

应该说明的是，上述问题只是领域专家所关心的比较抽象的问题。我们还将在本书适当的位置讨论更具体的问题。另外，在不同的知识获取实践中，知识工程师应该善于发现专家所关心的其他问题，并且考虑相应的面向专家的解决方案。

3. 关于本书

根据多年来的知识获取实践,我们清楚地认识到,知识获取的本质难点在于上面所讨论的心理切入、认知切入以及知识切入这三个问题。一旦知识工程师能够处理好这三个切入问题,知识获取也就进入了正确的轨道。这对后继的知识获取过程十分有利。本书着重分析知识获取的初期阶段,探讨知识工程师如何尽快地切入到专家领域问题以及专家知识中去。我们将不专门讨论知识获取后期的一些问题,如知识库的测试、求精、验证和验收。但是,我们不讨论这些问题并不等于本书所研究的内容与它们不相关。事实上,面向专家知识获取的一个重要目标就是减轻知识后期工作的压力。

本书共分为九章。前三章构成了本书的第一部分。它们讨论面向专家的知识获取的问题以及相应的解决策略。第一章站在领域专家的角度详细分析知识获取中的问题。这一章告诉我们,专家在知识获取中的作用并非是单纯的“问题解答器”。我们将在第二和第三章给出解决这些问题的策略。第二章讨论作为一般的知识工程师,他们应该掌握一个领域的那些重要的和抽象的方面。了解了这些方面后,知识工程师可以站在一个比较高的层次俯瞰专家的领域。第三章介绍若干具体的面向专家知识获取策略。这些策略都是从实践中总结出来的,它们具有较好的实用性。

第四章至第六章构成了本书的第二部分。它们讨论面向专家知识获取中双向了解,即知识工程师了解专家及其领域、领域专家以及其他领域人员了解专家系统。第四章分析知识工程师如何阅读领域书籍。在面向专家的知识获取中,阅读领域书籍被视为一项重要的知识获取任务。像其他的知

识获取一样,在面向专家知识获取中,知识工程师与领域专家的交谈是不可避免的。我们多次发现,知识工程师与领域专家的首次交谈尤为重要。第五章将分析知识工程师如何与领域专家进行首次交谈,并且讨论交谈中的注意事项。第六章讨论有关领域人员的概述报告。这种报告为知识工程师比较全面了解专家领域提供了帮助,同时也为知识工程师了解领域专家和其他人员提供了难得的机会。我们将分析在听这种报告的过程中知识工程师应该注意的重要方面。

在上述两部分的基础上,第七和第八章介绍知识工程师如何从领域人员那里获取领域知识。第七章介绍几种领域任务的分解方法。这些方法可以将原来的领域问题逐步分解,最后形成一个任务分解结构。分解结构为获取任务求解知识提供了天然的环境。第八章讨论如何为任务分解结构中的基本任务获取求解知识。第九章总结本书,并且列出面向专家的知识工程师应该具备的基本素质。

本书的阅读可分全面阅读和选择阅读。前者指的是从头到尾完整地阅读,而后者是指选择阅读书中的一部分内容。我们希望读者全面阅读本书。原因之一是笔者在安排书的内容时并没有刻意使得各章自成一体,虽然它们看上去各自独立。更重要的是,本书本身强调的一条知识获取原则就是尽量从整体上把握领域知识。

知识获取是一个很有前途的学科方向。1995年7月,笔者在北京召开的第二届青年科学家论坛上曾提出“知识高速公路”(knowledge highway)这一设想。人们常说,现代的社会是一个信息社会。其实,我们已经不知不觉地踏进了知识社会的门槛。在知识社会里,贫富这一概念将会在很大程度上失去其经济学上的意义,而越来越由人的知识数量和质量