

彼得.陈访谈录

本专访主要介绍彼得.陈的成功之路, ER 模型的起源及其在软件工程中的应用, ER 数据库的应用需求, 以及其他精彩内容。

玛丽安.温斯利

玛丽安: 欢迎来到ACM SIGMOD RECORD 数据库领域著名学者系列访谈。我是玛丽安.温斯利, 现在我们在圣迭戈, SIGMOD&PODS 2003的主会场。坐在我旁边的是彼得.陈, 他是路易斯安那州立大学计算机科学系的墨菲.福斯特教授。彼得以提出实体关系模型著称, 该模型被广泛用于企业数据建模。他在1976 TODS上发表的关于ER模型的论文已成为计算机科学引用率最高的文章之一, 被大多数大学教授评为最有影响力的论文。此外, 他还荣获了IEEE Harry Goode奖, ACM/AAAI Alan Newell奖, DAMA 国际成就奖, 以及 Stevens软件创新奖。他是IEEE, AAAS, 以及ACM会员, 在哈佛大学获得博士学位。现在, 欢迎彼得的到来。

彼得: 谢谢!

玛丽安: 彼得, 众所周知, 您是最著名数据库专家之一, 但是您的职业生涯与我以前采访的其他数据库专家有很大不同。您发现了一条不用写太多论文的成功之路! 您是如何取得如此巨大的成功? 更通俗地说, 什么使得您那么出名, 那么受人尊敬?

彼得: 正如你刚才提到的, 出名有很多不同方式。或者你写非常多论文, 或者你只写几篇论文, 但是每篇都要有很大影响力。但我认为最关键是你是否解决了一个真实的问题, 一个人们非常关注的问题。而我非常幸运地解决了至少一个这样的问题。

玛丽安: 很多年轻人想要搞清楚什么才算是真正的大问题, 您能给他们提些建议吗?

彼得: 我认为大多数人看世界和看问题的角度是不同的。但有时由于一些外部影响, 他们不能坚持自己的观点。因此, 我建议这些年青学者一旦工作确定或者开始任职, 他们就应该花更多精力到自己关注的问题上, 走自己的路。不仅因为他们应该发展自己的信念, 还因为他们应该遵从他们的直觉和信仰。坚定你的信念, 并为之奋斗不已。

玛丽安: 为什么ER模型一直那么受欢迎?

彼得: 我认为是因为很多人不知道怎么设计数据库, 怎么组织数据。ER模型为他们提供了一种很自然的组织信息方法。我们认为实体和关系的概念是最基本的概念。在我们看世界的时候, 我们看到了什么? 在这个采访的房间中, 我们看到一台摄影机, 人, 椅子, 是不是? 你坐在一张椅子上, 我们一同面对一个摄影机。这就是你, 我, 椅子, 摄像机之间的物理关系。同时, 在这个特定的时间, 我们都在同一个房间。这就是我们的时空关系。所以, 实体和关系是一种自然的组织事物和信息的方式。

玛丽安: 关于数据库研究, 我们应该关注什么问题?

彼得： 将来图形数据库会代替现在的文本型数据库，该数据库无论是用户接口还是内在结构都是基于图形的。

玛丽安： 凭直觉您认为下一个自然的数据库图形接口会是什么样子？我想现在还没有看到关于图形接口的正式设计准则。

彼得： 可能不需要用键盘输入文字，而是用个性化的图标作为输入。也就是说，你可以用一个图标代表一个人，不同的小图标代表不同的个体。这些就是图形接口；同样在底层也以同样的方式表示数据。

玛丽安： 让我们谈谈在ER概念细化之下的概念。ER图概念背后是什么语言概念？

彼得： ER概念与英语以及其他自然语言的结构非常吻合。名词常常对应于ER模型中的实体，动词对应于关系。另外，英语还有形容词和副词。形容词对应于ER模型中实体的属性。副词对应于关系的属性。由此可见，他们之间存在一个非常紧密的一一对应关系。

玛丽安： ER模型与古埃及象形文也有关系吗？

彼得： 是的，这就是我在过去的几年中花了很多时间看古埃及文的原因。实际上，我大概已经看了20本用古埃及文写的书。

玛丽安： 哇！

彼得： 我或许应该开一门古埃及语课程。实际上，埃及还发明了好几种类似于汉字的象形文字来表示现实世界。他们用不同的符号表示大自然中的不同事物和事件，以及不同实体之间的关系。非常有趣的是古埃及语和汉语有许多概念十分相似。有些文字，像“太阳”和“水”，是完全相同的符号。相隔千里的人们思考的方式竟完全相同，这一点非常有趣。与ER模型相似，两种语言不仅有塑造真实世界的规则，还有合并规则。所以你可以合并两个概念成为一个新概念。

玛丽安： 我听说有些中国方言是世界上最难学的语言，他们与ER模型也有关系吗？

彼得： 就发音来说，汉语可能是最难学的语言之一。

玛丽安： 那么书面语呢？

彼得： 我的意思是当你看到文字时，你不知道如何发音。但是有一些简单的规则帮助人们记住5万个单词，这与ER模型十分吻合。例如，有一些关于如何合并多个事物以及多个不同类型概念的规则。这些规则经常被用于信息技术领域中，在汉语中也是普遍存在的。

玛丽安： 将来有没有可能采用其他的语言概念来改进现在的概念建模技术？

彼得： 非常有可能。我一直都在研究这些可利用的概念，因为我相信一种成功的信息建模技术必须与人类的思考方式十分吻合。语言概念已经用了这么多年，也许当中有一些概念可

以借鉴。

玛丽安：关系模型直到现在都这么受欢迎。对于人们来说，表是一种很普遍存在的结构吗？这是不是关系数据库取得如此成功的原因？

彼得：是的。表是一种很普遍存在的概念。在很多地方可以看到表，你会看到很多诸如电子表格的矩阵。

玛丽安：古埃及文也用表吗？有发现过吗？

彼得：至今还没有发现。但是我所说的表只是其中一种普遍存在的结构，并不是唯一一种。关系模型为我们提供了一种特别的数据结构形式，但是我們也需要考虑其他形式。

玛丽安：其他形式是指到现在还一直在研究的面向对象模型或其他模型吗？

彼得：面向对象模型是一种不同于表的数据结构形式。但是，我认为实际上ER模型是在面向对象模型和关系模型之上的。因为ER模型更面向概念。而面向对象模型和关系模型更面向实现。所以在这两种模型之间做选择，就相当于在两种不同的实现方式之间做选择。

玛丽安：现在，在概念建模领域最主要的开放性问题的有哪些？

彼得：有很多很多问题。这就是我们每年都要开概念建模年会的原因（<http://www.cs.fudan.edu.cn/er2004/>）。其中一个开放性的问题是寻找一种以最自然的方式表示信息或数据结构的方法。是二元的还是多元的？是有向关系还是无向关系？哪一种结构依赖于语言，哪一种依赖于语言学，哪一种依赖于文化？也许有一种结构对于讲英语的人们来说是自然的，但对于其他人并不是如此。如若我们在外太空发现生命体，他们与我们的思考方式可能大不相同。

玛丽安：我确信非常有可能。以前有人提议在软件工程中使用ER模型。现在已经开始使用了，还是仍处在研究阶段？

彼得：有很多人已经开始在软件工程中使用ER模型。结果发现，在实际应用开发的过程中，最常使用的两个图就是ER图和数据流图。所以，在设计大型系统时，ER模型已经成为一种标准技术。

玛丽安：您对象UML这样的新型建模语言有什么看法？

彼得：UML模型在很多方面都表现不俗，但是它也有一定的局限性。优点是它能帮助实现ER概念，使得面向对象的程序员将精力投入到概念建模本身的重点上。我个人认为UML是一种语言，一种图约束。它可以帮助建立ER概念模型，但是他不能代替ER模型。认清这一点很重要，因为我认为UML像其他高级语言一样，像ALGOL，PL/1，FORTRAN等。这些语言中每一种都有控制结构，数据结构，部署等概念。几乎每一种编程语言都有这些东西。同样，UML是一种以面向对象的方式实现核心ER概念的语言，但是不论现在还是将来它都不会是唯一一种实现ER概念的语言。简而言之，ER概念是最基本的概念建模准则。几千年以来它一直伴随

着我们，乃至几十年以后它还将伴随我们。

玛丽安： ER方法如何帮助XML建模？

彼得： XML基本上是一种树形结构语言。研究XML的组织了解社会需要更灵活的结构，所以他们致力于开发新的模型来满足这些需求。我曾经受W3C XML SCHEMA和 Xlink工作组的邀请，以受邀专家的身份与这些工作组做过交流。如Xlink工作组，他们正在致力开发一种链接结构更加复杂的超链接。我们可以把一个超链接到更复杂结构的扩展看作关系概念从低级关系到高级关系的演变。这与操作系统中物理地址到逻辑地址的演变过程很相似。这也是从两点链接到多点链接的演变。

玛丽安： 所以，这与从M到N超链接很相似，它可能会有自己的属性吧？

彼得： 完全正确。

玛丽安： 在过去25年中，ER会议对社会产生哪些重大影响？

彼得： 首先，它为人们提供了一个交流和提出新想法的好场所。同时它还建立了一个新的研究领域。提供了一个让各种各样的人齐聚一堂的机会。参会者不仅仅是来自数据库领域的，还有来自其他领域，如信息管理系统、软件工程、人工智能以及信息检索领域。有时，人们会讨论ER模型在音乐制作和组织设计中应用。所以会议还提供了一个打破常规的途径。事实证明会议还促使概念建模领域成为一个倍受瞩目的研究领域。

玛丽安： ER建模对系统交互有什么帮助？

彼得： 我们已经开发了非常多的系统，但是大多数系统之间不能交互。这有很多原因，最重要的原因是它们的数据结构互不兼容。我们今天所拥有的系统就像是相互孤立的小岛，而我们真正需要做的是在小岛间建桥。为了这个目的，我们需要发现它们之间隐藏的实体关系结构。有时候这些结构被隐藏，有时是暗含其中。因此，多数情况需要数据挖掘技术发现实体间蕴含的关系。

玛丽安： 如何把结构建模扩展为功能建模？

彼得： 首先确定结构，接着在这个结构上加入功能，也就是操作部分。在工业界，人们通常分开做这两件事。在一个组里完成数据和信息建模，在另一个组里完成功能建模。这些模型可能互不兼容，这些建模的人也是不同的。我想在将来这两类建模都会由同一组人用相同的技术完成。这样我们就可以有一个既考虑数据因素又考虑功能因素的模型。

玛丽安： 结构设计师会使用像ER这样的方法吗？

彼得： 他们现在就在使用。在设计结构时使用ER或类ER技术，在设计功能时则使用不同的技术。

玛丽安： 在功能方面，人们现在使用什么类型的技术？

彼得： 现在人们使用数据流图，状态转移图，或其他技术。我建议将来在ER概念基础上增加一些东西使其成为统一的建模技术。到那时，我们会拥有一种可以同时完成结构建模和功能建模的技术。

玛丽安： 好吧，那就叫实体关系功能模型。

彼得： 好，你已经有了名字。

玛丽安： 好的，那您必须要写文章了！

彼得： 没问题。

玛丽安： 我记得90年代洛阿尔托研究中心曾经用过一个基于ER模型的数据库管理系统。既然ER模型那么流行，为什么ER数据库系统没有流行起来呢？

彼得： 首先要明白其中涉及到技术、政治、经济问题。主要原因是时机不对。过去人们有正确的技术思路，但是当时时机不对。在经济上，主要的问题是大量投资都集中在某些相对成熟的技术上。25年前，在IMS层次数据库技术上投入了大量的资金。你可能听说过Ted Codd博士说服IBM高层实现关系技术是多么困难。他曾经对IBM高层特别失望，言辞激烈。现在关系数据库系统技术成为主流技术，工业界对它的兴趣比起25年前大太多。所以由于外界的压力以及惰性，引入一种新的高端技术很困难。这也是面向对象数据库管理系统所面临的一部分问题。学术界讨论过很多有关数据库管理系统的问题，而商界由于在关系技术投入很大对此毫无兴趣。但是，我认为时机在变；ER型数据库管理系统的时代就快到来。我想10年后，你就会看到许多相关研究，原型系统，甚至是商用系统。

玛丽安： ER数据库管理系统和对象关系数据库系统有什么不同？

彼得： 对象关系系统只与对象相关。那么，什么是对象？只看“对象”这个词本身，它基本上是一个很抽象的概念；但是一旦它被实例化，它就是函数以及变量的一个封装。由于使用这个概念方式很特殊，有些函数很难理解。例如，在面向对象的系统中如何实现2+2？你需要发送一个“+2”的消息。

玛丽安： 这看起来有些难以理解，是不是？

彼得： 是的，很难懂。有些东西是不能用面向对象的方式很自然地表达出来。因此，会引起一些问题。从25年前直到现在，我的观点一直都是要以一种自然的方式做事情。任何不自然的东西都很难实现以及被工业界和公众所接受。

玛丽安： 所以，现在面向对象数据库仍是非自然的？

彼得： 对，但是，我认为与对象关系数据库管理系统相比面向对象数据库管理系统更加不自然。比起面向对象数据库，对象关系数据库更有市场价值。如果一个对象关系数据库能够避免非自然的特性，就没有什么可挑剔的了。

玛丽安： 您认为ER数据库将会在哪些应用领域中得到认可？

彼得： ER数据库将会应用到那些高层与概念很吻合的应用领域。例如，假设我们要发现恐怖主义者。谁可能会是恐怖分子？你可能知道是否有人买了化肥；是否有钱被转移到其他地方；这些都是复杂关系。为了给你提出的高级问题（“谁是恐怖分子”）提供一些可能的答案，系统会提出一些问题。这类查询用ER系统完成会非常自然。

玛丽安： 为什么说这比用纯关系系统更容易？

彼得： 在纯关系系统中，连接关系不太明显。同时，数据是分散的，你不知道什么和什么相对应。有时你会对应错。假设你知道人名和船名，但是放到一起可能会建立错误的连接。在ER系统中不存在这样的问题。你不会建立错误的连接。虽然我们有那么多信息，但没有时间关心是否存在不正确的关系。我们不能额外花时间去找到不正确的连接和不相关的数据类型。

玛丽安： 刚才您对识辨恐怖分子的技术很感兴趣。那么您对美国政府的国家信息意识工程所引发的隐私问题有什么看法？

彼得： 安全和隐私存在一种平衡，我们会发现我们身处一个奇怪的世界。例如，我住在路易斯安那州，在那有些全国都关注的问题，诸如连环杀手，连环狙击手等等。

玛丽安： 哇。我提议以后到新奥尔良举办SIGMOD会议吧。

彼得： 好呀，将来我们会在那里举办的。在过去的一年里，在路易斯安那州巴吞鲁日发生了许多大事件。过去，美国有线电视网或者今日美国会说“路易斯安那州，巴吞鲁日”，现在他们会去掉州名。因为现在巴吞鲁日出了名。如果我们有监测集成大量数据的能力，我们就很快找到连环杀人犯。但是事实上我们并没有这种能力，因此浪费警察大量的时间和精力，还使得许多人死于非命。据说几周前警察利用DNA抓到一名嫌疑犯。以前我曾经提议过建立一个DNA银行。但是，这是一个很敏感的政治问题，和你刚才所说的隐私和安全的问题直接相关。

所幸，我们不必使用任何引起争议的技术。因为，至今为止还没有开始全面推广这些合法可用的技术。例如，很多911恐怖分子都有像违规驾驶之类的犯罪记录；其中一个罪犯在案发前两天曾被马里兰的警察拘留过。但是警察不知道他就是CIA监控的对象，因此放他走了。另一个著名的911恐怖分子安得也曾在佛罗里达被拘留过，还被通缉了，但是警察并不知道这些。这些情况都不存在隐私问题；所有信息都是公开的。主要问题这些数据之间没有关联，没有建立起它们之间的关系，也许对全局把握不足。这使我们更加意识到利用可获得的数据，将数据关联起来，以及利用现有技术进行推导的重要性。当然在做这些事的时候，我们应该遵从法律，我想这不会引发隐私问题。

玛丽安： 假设您现在有充足的时间可以做另一项你没做过的研究，那么您想研究什么？

彼得： 我想做广义实体关系模型理论研究，使得它更加复杂可以应用到更多的应用领域。数据和结构可能有多层，有不同的代数操作符。正如我们之前讨论过的，集成功能建模和数

据建模对ER模型而言会是一个不错的发展方向。我们不仅有数学部分还有图形接口，这对模式设计师以及终端用户都会有很大帮助。

我还想将ER模型概念扩展到其他领域。不仅仅是信息技术领域，还有其他地方，像人际关系。你是如何管理你的人际关系的？在商界叫做客户关系管理。一旦人们开始关心并好好了解人与人之间的关系，那么他们会更好处理人际关系，他们会互相感激，他们之间的关系会密切。当你开始研究一个理论时，你会希望把这个理论应用到许多不同的领域当中。当你在不同的应用领域中做调查时，你会得到许多反馈，这对你改善你的理论结构以及其他事情会很有帮助。过去我一直从事WebDB的研究，我听过说乔·赫勒斯坦在讨论不同领域中结构挖掘的问题。这正是我想采用的方法，我想其他人也想这么做；云深不知处，只缘身在此山中。有时候跳出来看问题会更清晰，现在我们太关注数据处理了。看看其他领域，你可能会找到一个可以用到概念建模的功能建模中的新的方法。这些就是我想做的。

玛丽安：作为一名计算机科学家，假设您可以改变一件发生在您身上的事，你想改变什么？

彼得：没有一件！我不想改变任何事，因为我一直都非常幸运，非常幸福，上天赐予我很多好机会。有时我会想我真是太幸运了！2000年我在盐湖城的ER会议会议上见到了菲尔·伯恩斯坦，他说发现ER概念是一生一次的机遇，我非常同意。我只是在正确时间，正确的地点，提出正确的想法。可能在将来也不会发生，只是在当时那个时刻碰巧发生了。对于我而言，这是非常幸运的，非常幸福的，给了我好多机会。所以我不想改变任何事。我非常感激上天赋予我的一切，我非常高兴能有机会为促进技术发展做出贡献。

玛丽安：非常感谢。

彼得：谢谢。

(富丽贞译，范玉雷校)