

第2篇

吉奥·威德厚德访谈录

Interview with Gio Wiederhold





吉奥·威德厚德简介

吉奥·威德厚德 (Gio Wiederhold)，博士，斯坦福大学计算机科学系教授。1957 年在荷兰获得航空工程学士学位，1976 年于美国旧金山的加利福尼亚大学获得医学信息科学的博士学位。主要研究基于知识技术的应用和开发，并重点关注中间件和中介器等技术。他是 ACM 会士，IEEE 会士，Golden Core 成员，美国医学信息学院会员，SIGMOD 贡献奖的获得者，并曾是 ACM 数据库系统学报 (Transactions on Database Systems) 的主编。

本专访介绍了吉奥·威德厚德职业生涯中期进入学术界，怎样成为一位有效的顾问，为什么要在一个基金机构担任项目经理，本体代数、仿真和更多的需求。



玛丽安·温丝特

本期吉奥·威德厚德的采访是在 2001 年 7 月，是吉奥·威德厚德从斯坦福大学退休庆祝的前几天。吉奥已在斯坦福大学工作多年，活跃在计算机科学和医学信息领域。20 世纪 80 年代末，吉奥做过几年 DARPA 项目主管，主要关注中间件和中介器。吉奥是 ACM 会士，IEEE 会士，Golden Core 成员，美国医学信息学院会员，SIGMOD 贡献奖的获得者，还担任过 ACM 数据库系统学报（Transactions on Database Systems）的主编等。

吉奥的职业生涯多样有趣，他在世界各地工作期间的奇闻轶事和冒险活动数不胜数。不过有关他的奇遇故事还是大家自己去问吧。吉奥很喜爱讲故事，说笑话。

玛丽安·温丝特：这次访谈是 SIGMOD Record 一个新特色专栏，我想在您退休之际采访是一个能让您迸发名言警句的大好机会。

吉奥·威德厚德：你们还采访过谁？

玛丽安·温丝特：噢，还没有。您是首位受访者，这意味着我的采访还不熟练。

吉奥·威德厚德：好啊，你可以用我来做试验品。

玛丽安·温丝特：吉奥，当您在四十多岁的时候，已有稳定的事业，有一个孩子，即将有第二个孩子，这时候您放弃了计算机系统经理的事业，而去攻读博士学位。显然，这不是一件容易的事，是什么促使您这样做的？

吉奥·威德厚德：之前我的工作从技术工作转到管理工作上。管理工作是一种重复劳动。我要不断跟人打交道，而在过去的几千年人类并没有多大的进化。人们有很多的问题，像不爱工作，爱做错事，爱无休止地争论等，所以我考虑我更喜欢跟机器打交道。

玛丽安·温丝特：攻读博士学位是与机器打交道的唯一途径吗？

吉奥·威德厚德：是的，更多的投入到研究中，既可以提升技术水平又能远离管理。它也有点儿尝试的意味，因为当时我没有任何学位。

玛丽安·温丝特：您不是有一个荷兰的航空工程学位吗？

吉奥·威德厚德：是的，但那大约只相当于副学士学位的水平。

玛丽安·温丝特：我明白了。所以，您就直接从您的副学士学位跃升到博士学位。

吉奥·威德厚德：这源于我有很好的推荐信，如约书亚·莱德伯格等人的推荐。如果你从诺贝尔经济学奖获得者那里得到推荐，会对博士录取起到很大的作用。

玛丽安·温丝特：现在，您已经在斯坦福大学计算机系做教授 20 多年了。您能讲述一下您在那里的工作吗？作为教授 ...

吉奥·威德厚德：我的任职在计算机科学和医学之间来回转换，有时全职在计算机系，有时在计算机和医学系同时兼职。其原因是斯坦福大学的研

究型教授（Research Professor）名额有限，大约只有 10% 的教师可以从事研究型教授工作^①。斯坦福大学的计算机科学系大约有 26 位教职员工，当时布鲁斯·布坎南和汤姆宾福德在做研究型教授。我的博士学位也被医学信息学认可，且当时我的医疗计算的论文数量与计算机科学的论文数量相当。医学计算一直是一个很能激发人们的应用领域。

玛丽安·温丝特：哪件工作是您觉得最值得骄傲的？

吉奥·威德厚德：应该是在获取终身教职之前做的事情：我为斯坦福大学医学院开发的 ACME 分时系统。

玛丽安·温丝特：当时 ACME 系统的特别之处是什么？

吉奥·威德厚德：它的集成很好，所以，例如编辑器只是编译器的一个函数。这是一个单语言系统，我们实现了 PL/1 语言的一个子集，结果发现很受益。该技术本身是很有趣的：它是一个增量编译器，其中没有很多实例存在，但它很好地平衡了传统的编译器和解释器。它的工作方式是，PL/1 的每一行，无论它是一个语句还是多个小语句，都被编译，随后插入到链表结构中。因此，没有跨语句优化，但每个语句都被完全编译。之后，当用户想要改变一个声明，我们只需要简单地改变链表结构——这样我们就可以插入语句，删除语句。用户在一个程序中唯一不能更改的是类型。例如，用户不能将数值更改为字符串，因为太多的代码将需要重新编译。

ACME 编译器的灵活性是系统的实时性方面的非常重要的问题。该系统做了实时数据采集。当程序出现问题时，ACME 系统允许用户更改程序而医疗实验仍然进行着，不会取消实验（往往包含动物等）或丢失数据。

① 美国许多大学有两类教授职位：终身教职型（tenure-track）和研究型（research）教授。两者的主要差别在于：一旦获得终身教职，其职位一般不能被解除，除非有特殊的情况发生；其薪水由大学支付，相应地，他有义务完成一定教学工作量和系里的事务性工作。研究型教授的薪水则来自其科研经费；其与大学每五年续签一次合约；通常他们有很轻的教学工作和事务性工作，甚至不必承担任何此类工作。

玛丽安·温丝特：您是什么时候开发 ACME 系统的？

吉奥·威德厚德：我是 1965 年末开始，1967 年完成。我们当时买了大量的内存，差不多一兆字节的内存，这在当时是非常庞大了，我记得当时从波基普西来的卡车司机，每天晚上都会打电话给我，告诉我他已经载着那块巨大的内存走了多远。

玛丽安·温丝特：ACME 的用户知道他们在使用着什么样的系统吗？它的一些功能在今天来看仍是新颖的，放在当时就更不用说了……

吉奥·威德厚德：这是他们中大多数人使用过的第一台电脑！后来，事实上，我遇到了用过更新更好的系统的医务人员，他们为程序中出现的错误导致丢失数据，并不得不重新开始他们的实验而感到沮丧。

玛丽安·温丝特：我听过一些人说，在数据库研究领域，好的技术都已在工业界实现了。对此您怎么看？

吉奥·威德厚德：不，我不认为这是真的。数据库核心技术确实已经成熟，但应用中的问题还是层出不穷的。随着我们对应用研究的深入，还有许多需要研究的技术。比如，我一直专注于信息集成的一个问题：语义差异，其重要性我早期并没有意识到。长期以来，我都是在研究模式集成。我在咨询中了解到：人们先做模式集成，但集成起来的数据库依然不能一起工作，因为术语在不同的数据库中含义不同，作用域也不同，等等——所以我最近在研究本体代数。

玛丽安·温丝特：那是什么？

吉奥·威德厚德：为了集成不同来源的数据库或数据，你必须解决本体的差异。我不认为集成只是使数据库变得越来越大，如 UNION 操作那样。其实集成是要使数据关联起来，对特定的应用便于找出其连接数据何在。这方面的一个简单的例子是鞋厂和鞋店关联。你不需要集成来自鞋厂和鞋店数据库

的所有部分，但为了能相互合作，必须在两个数据库中交集上达成模式匹配，如鞋子的尺码、款式、名称等鞋店买家能够理解的有关信息；但不必将鞋店与鞋厂的人事信息集成在一起。事实上，只是基于简单的术语匹配来做集成是大错特错的——就像鞋店里的指甲（nail）是消费者身体的一部分，而在鞋厂中它（nail，钉子）是用来钉鞋跟的。因此，这些简单的匹配是行不通的。

此外，当你从不同的来源获取数据时，也会有粒度的差异。例如当我在家里与我的儿子一起工作时，我们需要一些钉子，我会告诉他，从咖啡罐里拿一个长的，头很大的钉子。而木匠会使用像“箱钉”这样的术语。木匠知道很多钉子的名字，像爱斯基摩人对冰的命名一样多。但让我学习所有木匠的术语是不可行的。反之，要木匠使用我的术语也是不可行的。

所有这些类型的本体转换可以形式化，而这也正是可以使用交集、投影和类似操作的地方。但是这些转换必须是基于知识的操作，因为它们包含匹配规则。我认为用来处理本体的差异的方法提供了非常灵活的可扩展性，其中数据源可以保持自主和高效。

玛丽安·温丝特：您有关于通过本体转换处理语义差异的文章吗？

吉奥·威德厚德：有一些关于这方面的早期论文。Jan Jannink 早期的毕业论文是关于这方面的，Prasenjit Mitra 做了更多的形式化的工作。但在一定程度上，由于这类研究涉及到语义，因而可以很容易地研究上许多年。这些问题不是针对很传统的数据库应用，但对决策支持一类的数据库应用确实十分必要的，对其他一些数据库应用也是相当有用的。

玛丽安·温丝特：本体代数的想法是来自您的咨询工作么？

吉奥·威德厚德：是的，因为我写了很多东西，人们也试图按照我写的去做，因此我常被当做咨询顾问，在为人们咨询过程中我也发现我写的东西有些是不对的。

玛丽安·温丝特：您是否认为所有的数据库专家都应该出去做咨询吗？

吉奥·威德厚德：如果他们有能力做咨询工作的话，可以抽时间做做。咨询要求非常善于倾听。我的咨询秘诀是，花一整天时间倾听，然后想清楚你听到的东西。第二天提出问题并继续倾听。仅在第三天才开始提供建议。这种秘诀会让很多人入门时已经知道怎样做正确的事情。有时，在咨询中了解这一秘诀是有用的，但我不认为这是最有效的咨询，这种秘诀可能使你在咨询中学到的知识更少。当然，我已经通过咨询学到了很多，并且我也仍然在做。

玛丽安·温丝特：吉奥，作为预测未来的预言家，您有令人印象深刻的预测。中介器是这些预言中最知名的。您现在预测一下什么领域将变得重要？

吉奥·威德厚德：目前还不存在，但可能变成和数据库一样大的领域，就是数据库和模拟结果的结合，即使模拟和电子表格一样简单。目前我们说，用数据库做决策，而事实上数据库仅给了我们一个过去的历史。决策者不得预知未来。决策者需要把数据库技术和规划未来的方法相结合的信息系统，这可以从规划技术中借鉴一些东西。如今，决策者做的绝大多数的规划是非常缺少数据的，因此，这种规划常常是很不准确和很难修改的。正在完成的规划也是非常不灵活的。因此，我认为如果我们真的想拥有满足决策者需求的信息系统，我们需要提供无缝地穿越于过去和未来的技术。但是，这样的研究需要新的洞察力。这其中存在一些非常困难的技术，实际上，玛丽安，你应该对此有清楚的了解，因为你是研究这一领域的。

玛丽安·温丝特：是的，管理多优化方案是困难的，但从理论上讲已是过去。

吉奥·威德厚德：除非你具有历史修正主义。

玛丽安·温丝特：我记得您在 80 年代中期谈论过同样的研究课题。什么时候能够实现呢？为了中介器，您最终去了 DARPA（美国国防部先进研究项目局），并最终做了出来。为模拟这件事您想再来一次吗？

吉奥·威德厚德：我没有精力再次去 DARPA 并推进它了。但是我确实看到了对此类研究的需求。如果我和决策层的人们沟通，他们肯定能够理解这一问题。在我退休后，我将不能完全闲下来……

玛丽安·温丝特：我觉得您根本就闲不住！

吉奥·威德厚德：我会继续推动这方面的事情。但是，我希望其他人加入这一研究，因为在我的有生之年，我可能看不到最终的研究成果。

玛丽安·温丝特：这涉及到另一个问题。我听说目前 NSF 有 87 个项目经理的职位虚位以待。作为前任 DARPA 的项目经理，您为什么认为人们应该珍视这样的机会？

吉奥·威德厚德：DARPA 目前也有很多开放的职位，我们学者的数量还是不足。

如果大家把决定未来计算机科学研究资助的工作都留给官员去做，那么结果可想而知。官员往往具有伤害性，他们容易犯错，又不愿冒风险。因此，官员不适合作为项目经理。据我所知，成为项目经理是很难的，尤其当你有很多家人在那里工作的话，你不能随意地做事情。做一名兼职的项目经理也不是行之有效的方式。但是，做一名项目经理是可以在一定程度上改变世界的方式，并且可以改变事情发展的方向。要做到这一点，你必须心甘情愿地为此去做，并甘愿冒险为其奋斗。

作为学者，我们具有一点优势，就是做完项目经理之后，还可以回来再做老本行。我发现作为项目经理的最大负面影响是重返科研后需要花费大量时间重新开始。是的，我可以得到一些资助来重新开始我的研究项目，但是我得重新开始找学生。通常，研究需要一个连续的学生流，经费和学生两者对研究而言不可或缺。重启我的研究要花费差不多一年的时间，这远超出我的预期。

玛丽安·温丝特：您认为目前在技术方面最令人兴奋的发展趋势是什么？为什么？

吉奥·威德厚德：很明显，目前获取信息的方式更为广泛。这使得我们前边谈到的本体代数大有用武之地，因为信息来源具有更强的自治性。我们可以依靠 XML 表示一种通用的信息表述方式，但语义的问题仍未解决。

一个令人兴奋的发展趋势可能是在线发布，这也带来了一些非常难的问题。比如，在电子商务领域，怎样设定质量的度量标准。目前，所有互联网购物的比较服务是基于价格比对，但是对于我们购买的很多东西，价格往往是第二个考虑的因素。对于类似于书籍这样的可替代商品，主要问题是价格。但是当我买一台电脑或者一辆汽车的时候，价格仅是一个因素，也许质量、维护和可用性等等是更重要的因素。如果我要买一台投影仪，对我来讲，一个重要的因素是这台投影仪会产生多大的噪音。然而，当下我不能通过网络找出投影仪的不同噪音水平的比较数据。当然，噪音水平仅是测量的一个质量标准。把所有事情都归结为一个价格比对，有点太过简化我们的价值系统。

玛丽安·温丝特：对于终身教职或整个高校系统，您希望看到怎样的改变？

吉奥·威德厚德：我不是能太情愿地接受终身教职。事实上，我不是一位足够好的老师，因为我得不到好的教学反馈。在很多方面，一个非终身教职的职位更好。我可以对我不能教好的课程说不，因为，我对这些课程没有兴趣，或者是我对教课的方式没有兴趣。我记得有一门课名为计算机体系结构，结果教的是汇编语言，用了很糟糕的机器。你不会对这种课题有任何的激情。

我认为终身教授没有像其说的那样重要，尤其是研究型大学，因为你自己的薪水仅是支持你开展研究工作的一部分。我认为大学可以没有终身教职制度。对此我不置可否。

玛丽安·温丝特：您对数据库领域的初始或职业中期的研究者和开发者有什么建议？您是两者兼具。

吉奥·威德厚德：一个非常实际的原因促使我进入数据库领域。当开发 ACME 系统时，我搭建了一个实时获得数据的系统，收集到很多人们不能很

好管理的数据。因此我用到了数据库技术，并根据手头上的问题我稍微做了些扩展，构建了有关系统。对于新上路的研究者或从业者，因为数据库是一个很大的领域，我建议从做应用开始。以我为例，我是从医药和军队的应用开始的。这些应用看起来很难。在这两个应用中，你必须在决策者犹豫不决时快速做出响应。这种情况与商业领域的应用大有不同。在商业领域中，决策者从来不需要直接接触电脑，因为他们有准备报告和评估的中间经理。

无论你选择什么应用，需要自愿花一些时间深入理解这个应用领域。

对于开发者，我建议应学习足够的理论知识，并时刻关注领域内的进展，多读多看。我能进入学术界的原因大概在于：作为一名开发者，我每年一次将我做的事情记录成文——从我很早的项目开始，比如调查从 50 年代后期开始的火箭燃料燃烧情况等。那些不是最好的论文，他们在很多方面仅是我做事情的记录，包含一些定量的信息。

当你在工业界工作时，不必试图模仿撰写学术论文。应该从工业界做的事情中获益，你可以报告一些定量数据和计算的规模。这些对于学术界来讲是很有价值的。这比你写一些三流的学术论文更受青睐。

玛丽安·温丝特：在程序委员会中，您是否看到学术论文和工业界论文之间的差别？

吉奥·威德厚德：是的，我常常试图给作者一些反馈意见：你有很好的描述，也给了一些你的实验结果，但这不是使你取得成功的潜在哲理。

玛丽安·温丝特：您作为一位计算机科学的研究者，如果您可以改变一件事情，会是什么？

吉奥·威德厚德：我主要是理论的应用者；我可以恰当地应用理论。我希望能在理论上变得更强。幸运的是，当我的研究遇到理论问题时，我可以发挥优秀的学生和同事们的能力来解决问题。

（李勇，王璐 译，孟小峰 审校）