

第1篇

杰弗里·乌尔曼访谈录

Interview with Jeffrey D. Ullman





杰弗里·乌尔曼简介

杰弗里·乌尔曼 (Jeffrey D. Ullman), 1942 年生, 1963 年于哥伦比亚大学获得工程数学学士学位, 1966 年于普林斯顿大学获得电子工程博士学位。1966 年至 1969 年在贝尔实验室工作, 1969 年至 1979 年在普林斯顿大学先后任副教授、教授, 1979 年至今在斯坦福大学任教教授。1995 年, 他成为美国计算机协会 (ACM) 的会士, 2000 年被授予 Knuth 奖, 他还和 John Hopcroft 一起获得 2010 年 IEEE 颁发的冯·诺依曼奖章。研究兴趣主要包括数据库理论、数据库集成、数据挖掘和信息技术在教育中的应用研究。担任 Computer Languages (1974 ~ 1981)、Computer and System Sciences (1974 ~ 至今) 和 Theoretical Computer Science (1974 ~ 至今) 等多个学术期刊编委。发表论文 180 多篇, 出版著作 19 部, 其中包括《Database Systems: The Complete Book》(Prentice Hall, 2009)、《Elements of ML Programming》(Prentice Hall, 1998)、《A First Course in Database Systems》(with J. Widom, Prentice-Hall, 2008)、《Introduction To Automata and Language Theory》(Prentice Hall, 2006)、《Compilers: Principles, Techniques, and Tools》(Addison-Wesley, 2006)、《Mining of Massive Datasets》(Cambridge University Press, 2012) 等多部经典著作, 被视为业界的规范。因工作成果突出, 先后获得 Guggenheim Fellowship (1988 ~ 1989)、1996 年 SIGMOD 贡献奖、1998 年 Karl V. Karstrom 杰出教育家奖和 2006 年 E. F. Codd 创新奖, 美国国家工程院院士。

本访谈主要介绍了杰弗里·乌尔曼关于高等教育的未来、创业、数据库理论等方面的看法和建议。

玛丽安·温丝特



欢迎来到本期 ACM SIGMOD Record 数据库领域杰出人物访谈。其实我本身不太喜欢外出旅行，但是为了能够在每一期中都给大家呈现一个新的访谈，我还是决定离开位于伊利诺伊州的家。6 月份的时候我到加利福尼亚参加了吉奥·威德厚德（Gio Wiederhold）教授从斯坦福大学的退休庆祝活动。我在那儿采访了吉奥和他的两位同事杰弗里·乌尔曼（Jeff Ullman）和赫克托·加西亚·莫利纳（Hector Garcia-Molina），所以在后面的几期访谈中将会看到我和这三位来自斯坦福大学教授的访谈。

为了使访谈更加自由和轻松，我们在采访的时候进行了录像，随后再根据录像整理成文。这样我们就可以直接把录像放到 SIGMOD 网站上，大家可以直接听到这些大师都说了些什么。如果有录像的话，我强烈建议大家直接观看录像，因为我发现在我整理成文字的时候，有些细节信息还是遗漏了。

我发现根据录像整理成文也是一件极具挑战性的工作。我把访谈中的一些语气词、部分句子片段都翻译了出来，但是整理成文以后却发现这些信息带来了一些负面的效果。为了便于阅读，我对部分问题进行了重新排序，同时对长度也做了一些适当的调整。在整理成文的过程中肯定有很多错误，对此我向读者和采访对象深表歉意。同时我也建议大家最好通过录像来了解相应内容的准确含义。

第一个采访对象是 ACM SIGMOD 奖项的获得者杰弗里·乌尔曼。杰弗里在斯坦福大学计算机科学系工作多年。大家都知道，杰弗里非常有名，主要是因为他对数据库理论和计算机教育带来的巨大贡献。我之所以首先发表杰弗里的访谈，主要是因为他对初创公司以及这些公司对博士的影响做了一些非常及时的有价值的评价。

玛丽安·温丝特：珍妮弗·威多姆（Jennifer Widom）曾经告诉我您对在线教育有独特的见解，您能给我们谈一下这方面的事情吗？

杰弗里·乌尔曼：我觉得情况应该是这样的，大学教育需要老师的广泛参与，一对一地或者帮一小组人调试程序没有太大的意义。毫无疑问，计算机科学是一个一对一的过程，大学教育是教导学生如何认识人生以及如何应对整个世界，这是住宿式大学教育的一个优点。其实，自 13 世纪以来，大学教师就没有提高自身的教学生产力。我想做的是能按应用服务提供商（SAP）模式，提供一些在集中式教育环境开展得比较好的大学教育内容，同时和本地的老师合作，为少数人提供高质量的教学服务。同时我认为教学应该是集中式的。

很多开放型的大学和类似的教育组织机构都创建了自己的课程包，他们经常会这样说：如果你想了解数据库系统方面的知识，这就是所有人都需要知道的东西。我不敢太武断，但是很显然，我也希望所有人都选用我的教材来学习。我们认为不同的老师在教学内容的选择上会有不同的观点和看法。教师应该能够自由地选择相关课程内容和教学资料，同时也可以自由地调整讲授顺序。给学生讲授的内容随机性也不能太强，事先需要确定一个提纲。但是一般情况下我建议所有的数据库老师都选择 30 ~ 40 学时作为课程内容，这样就可以避免很多麻烦，他们可以不用再费事去准备这些讲义，也不用准备笔记、把笔记上网或者制作幻灯片。这些任务都可以一次性完成，节省大量时间。

我认为提供 24 × 7 的全天候的在线辅导对于数据库教学以及所有的计算

机课程教学都是至关重要的。作为学生，当你遇到问题的时候，你应该能够随时发送邮件求助，并且在世界其他地方有助教阅读你的邮件并立即给你答案，或者尽快给你解答。尤其是当人们学习使用一个系统或者是在调试程序，这种辅导方式就显得非常有效。这样就不用像很多其他同学一样选择放弃，也不用等到下次课或答疑时间，他们只需要等待几分钟就可以解决问题，然后继续学习。

在数据库课程教学中，我们发现一个非常重要的事情，就是培养学生用真实的 SQL 系统做一个实际的项目。学生自己选择项目，通过数据库设计并转换成关系模式，做各种各样的编程工作，包括利用 C 语言进行嵌入式 SQL 编程。这些经历非常重要，很多学校都认为维护一个实际的数据库系统供所有的学生使用非常困难，当然在斯坦福大学肯定可以提供这些服务。有时候我们好不容易找到了一个数据库管理员，但他很快又离职了。由此看来，数据库系统还是需要采用集中式服务的。我最近了解到 IBM 一直在为一些学校做这些事情。我希望这种服务的范围能够进一步扩大，最终能够为所有的学生提供服务。

玛丽安·温丝特：您认为 SIGMOD 或者斯坦福在做这些事吗？

杰弗里·乌尔曼：我们称之为 145.com（145 是我们的数据库导论课程编号）。我最终想做的是创建一个公司，公司最初的目的是为数据库教学提供服务，进而扩展提供所有的计算机课程服务，至少为其中的一部分计算机课程提供服务。我并不想为导论类的课程提供类似的服务，因为课程的层次越高，集中式的服务就越困难，因为高层次的课程一般都是研究性质的或者比较特殊的，内容变化相对比较频繁。我真正想做的就是想按照这种方式来处理所有的课程。有很多学校根本请不起数据库老师，甚至有些计算机系也请不到人。那么我们这种集中式的服务对这些学校就非常适合，同时对于那些有计算机系的学校来讲，如果他们需要开设一些相关的计算机课程，但是自己又开不了，这时候我们的服务对这些院系就非常重要。

当然，我们这种服务在有些情况下也是行不通的，比如历史、英语和心理学等课程，这些课程往往都有大量的老师，所以根本就不需要我们的服务。当然也可以做这样一个有趣的实验，看看这些学校是否认为下面的做法比较有效：辞掉三分之二的教师，把课堂教学外包出去，让剩下的三分之一的人来做所有的事情。

玛丽安·温丝特：关于这项工作，您有什么具体的时间表吗？

杰弗里·乌尔曼：我不知道，按照我的时间表，工作已经完成了！明年我希望能做一个实验。同时我也在设法找一些数据库厂商来一起合作。如果当这些工作完成的话，那么我就有更多的东西可谈啦。我们也正在和一些有兴趣提供帮助的大公司进行交流。

玛丽安·温丝特：您认为这仅仅是美国的事情还是全世界的事情？

杰弗里·乌尔曼：我认为这是全世界的事情。我想你可能会按语言来划分，当然有很多地方是用英语进行教学或者部分课程采用英语教学。我想把这些都融合进来。实际上，在国外我们确实也有一些朋友愿意做这件事情。

玛丽安·温丝特：关于课程收费问题，您有什么考虑或想法吗？比如以数据库导论课程为例。

杰弗里·乌尔曼：我做了一个粗略的计算，比如一个学生 200 美元。当然这仅仅是一个大致的计算方法。

玛丽安·温丝特：我听别人说您曾经说过“理论已经过时了。”您能详细解释一下吗？

杰弗里·乌尔曼：理论真的过时了。（微笑）

我想那并不是我真正要表达的意思。这要取决于我的谈话对象。如果我面对的是一个理论家，我会那样说；但是如果我面对的是一个专业人士，我

会强调理论远比他们想象的要重要。我认为这两种说法都是正确的。

20 世纪 60 年代末我们仅仅在 Don Knuth 的 LR(k) 语法分析器方面做了一些工作，其中用到了很多深奥的理论，当时有很多人都在努力解决这些理论问题。1965 年，Knuth 的论文发表以后，大量深奥理论的进一步发展促使了 yacc 的产生。yacc 出现以后，像约翰·巴库期及其团队不知要花费多少年开发一个 Fortran 编译器的事情变得相当容易，可能一下午就搞定了。这确是理论工作的一个巨大的成功，我们确实总是认为一切问题最终都会在理论面前屈服。实际上并不是这样的。（叹气）

除此之外，我还担心一些按照这种行为方式做事的人。我们在开始的时候往往都是由导师指定几篇论文，然后看能不能改进一些，我们往往就会说可以改进。接下来就会对文章中的一些假设做一些简单的修改，从来不考虑这些修改有没有什么实际的意义。当我们把问题做了一些调整，有了一些新的想法、算法和定理，然后就发表了另外一篇文章。五年以后，又有一些别的学生在我们的论文的基础上做进一步的修改。最后整个问题与实际问题的相差甚远了。

从某种意义上来说，我们应该很庆幸计算机科学具有抽象思考的能力，而不是仅仅以具体事物为基础。这种能力使得计算机科学成为一门令人兴奋的科学，同时也给我们带来了大量强大的工具集，但是同时我认为这也有巨大的风险。你会经常发现在一些理论研究团体里面，他们仅仅解决一些他们自认为有趣的问题，并假定他们所做的工作是有实际意义的，我一直为此而感到内疚。另外，如果理论不能解决我们想要解决的问题，那么这些理论很容易被忽视。而实际上，很多理论概念已经悄悄地进入了我们的头脑，并且在无形中帮助我们分析和解决问题。举例来说，我们曾经在依赖理论、有损连接和多值依赖等方面做了大量的工作。我们原本以为这些理论研究可以帮助我们实现模式的自动化设计或者类似的工作，但是事实上我们却从来没有达到过这个目标。但是每一个进行数据库设计的人员都必须了解什么是函数依赖，也必须知道当进行规范化操作的同时会损失一些其他的什么东西。如果

不掌握背后的一些理论知识，即使是一些专门人士也弄不明白。这就是我为什么在数据库领域总是要求人们要更加认真对待理论，至少要比八十年代更加重视。

玛丽安·温丝特：您还曾经说过，“一个成功的商业计划应该可以作为一个博士论文”。

杰弗里·乌尔曼：我不知道什么是成功的商业计划，我只知道有很多成功的商业。

这里我想表达的是，创业与过去已大不相同，现如今创业如同科研调研。当下很多大公司没有自己内部研究部门。像 IBM 这样的大公司有大量的基础研究计划，他们从中受益匪浅。朗讯、贝尔实验室（贝尔实验室也曾是朗讯的一部分），内部都有大量的基础研究项目。微软在基础研究方面也投入了大量资金，但是也有很多大公司（如甲骨文、思科等）确实没有自己的基础研究项目。但是，这些公司往往需要吸收很多新的创意，这样他们必须支付大量的资金，从那些拥有好想法的小公司来购买这些创意。

我一直都认为能够证明你的创意有效性的最好办法之一就是创办一家公司，利用你的创意做出一些人们需要的东西。我认为，学生应该拥有更多的自由，应该利用自己的创意去获得风险资金，创办公司，然后再卖给别的需要你创意的一些大公司，而不应该仅仅是把自己的创意发表在一些期刊杂志上，根本就没有人会去读这些文章。我想强调的是，我们已经经历过虚假繁荣和经济大萧条，但是长远的发展趋势应该是：越来越多的学生愿意基于自己论文中的研究工作创办自己的公司。越来越多好的创意应通过初创公司展示出来，而不是以论文的形式发表在期刊上，等着十年以后有人读到它并做一些与之相关的研究工作。

玛丽安·温丝特：那么博士学位是不是也就不那么重要了昵？

杰弗里·乌尔曼：不，不，我觉得博士学位还是能代表很多东西的：它

表明你已经学会如何做研究，如何进行创新，如何选择一些值得研究的问题，如何基于这些问题推导出该研究领域已知的一些知识，如何发现一些本来应该已经被发现、但是至今仍未知的一些知识，以及如何把这些知识结合在一起。

三个工商管理硕士（MBA）利用互联网实现了一个非常普通的商业策划，并创立了 archetypical.com 公司。以前我们常开玩笑说，这不是一个计算机博士应该做的事情，但是现在情况却并非如此。现在很多有价值的技术创意已经通过初创公司进入了市场。我以前在 Jungle 的几个学生令我感到非常自豪，当时我们正在做信息集成方面的一些研究工作，他们就想到可以在 Web 上做同样的事情。当然他们是幸运的，他们的时机也很好。同时重要的是他们确实拥有比较好的创意和技术，并且是第一个做这件事情的人。另外一个成功的案例就是 Google：当时他们确实有一到两个重要的想法能够使他们的搜索结果比其他的都好。如果当时他们仅仅是就此写一篇博士论文，再找几个本科生做一个简单的测试，看他们是喜欢 Altavista 的结果还是喜欢 Google 的结果，如果是这样的话，确实是有点愚蠢。这虽然并不能说明什么问题，但是他们确实迈出了第一步，并创办了自己的公司。我想在 Google 出现之前，Inktomi 是最大的搜索服务提供商，而现在人人都在使用 Google 搜索。这就表明，他们的创意确实有效，并且这要比去证明一些愚蠢的理论或是找一些本科生做一些简单的实验要好得多。

玛丽安·温丝特：几年前，如果我要招聘数据库新教师的话，您可能会告诉我您不知道一个博士毕业生为什么一定要进入学术界。您现在还这样认为吗？您是否认为他们都应该到初创公司中去？

杰弗里·乌尔曼：我可能会这样回答你：从长远来看，创办公司的经历会帮助他们成为一个更好的教师。毫无疑问，一些风险投资家在一些毫无意义的项目上浪费了数百万美元，而真正比较好的创意现在却很难得到资金支持。并且我发现，现在学术环境比以前更加规范了。但是我想说，可能之前

我和你谈论招聘问题的时候已经说过，现在当一个教师要比以前难多了。以前政府认为计算机教师应该有足够的资金开展研究工作，因为培养更多的计算机科学家符合国家的利益。现在培养计算机科学家同样也符合国家的利益，但是我们不能忽略在 20 世纪 60 年代和 70 年代由于 IT 人才的短缺，国家才愿意为计算机教师提供足够的资金支持。现在政府却觉得计算机专业不再需要一流的教师了。当然，关于一些研究型大学的情况，我想你也清楚是怎么回事。

玛丽安·温丝特：我确实了解，当然我们的读者不一定清楚。

杰弗里·乌尔曼：正好你现在正在整理稿件，你可以把这些加进去。如果一个学生有机会获得一个企业的研究员职位或者可以进入一家初创公司的话，那么你很难说服他去当一名教师。

玛丽安·温丝特：我们现在还是坐在您位于斯坦福大学的办公室里进行谈话，那么为什么您还在这里呢？

杰弗里·乌尔曼：我正准备退休呢。

玛丽安·温丝特：然后创办一家公司？

杰弗里·乌尔曼：不，我准备随便逛逛，做点事情，写点书，做点咨询工作。我年龄太大了，办不了公司了。

玛丽安·温丝特：创办公司一定得比较年轻吗？

杰弗里·乌尔曼：确实是这样。要想成功创办一家公司需要投入很多的精力。我看到一些年轻人满怀激情地创业，我想我年轻的时候也能像他们一样付出自己的努力去创业，但是现在肯定是不行了，精力差远啦。

玛丽安·温丝特：其实我也有过创业的经历，结果是没有成功。关于工业界的实验室您没有谈论太多，实验室还有吸引力吗？

杰弗里·乌尔曼：企业实验室确实有很多不错的工作机会。从人员比例上来看，企业实验室为计算机博士提供的研究职位和过去基本上一样多。但是，从人数上来看，与二、三十年前相比，在企业实验室工作的人却少了很多。我发现学生非常看重微软、IBM Almaden 的工作机会。我们还有一些学生去了朗讯。其他的地方去的似乎很少。还有一些学生认为其他任何地方都不如施乐帕洛阿尔托研究中心（Xerox PARC），虽然我不确定那儿到底有没有前途。当然，我可以肯定的是工业界还有很多其他比较好的工作机会。毫无疑问，工业界认为没有必要在基础研究上进行过多投资。众所周知，基础研究很难有所回报。当你在基础研究上进行投入的时候，你实际上是在为公司营造一种科研氛围。当年我博士毕业进入贝尔实验室的时候他们就是这样做的。他们根本不在乎我两年之中做了什么事情，他仅仅是想让人们知道贝尔实验室可以得到任何他们想要的人才。很多人在进行了若干年的基础研究之后，最终成了高级研发的领导者，这仅仅对贝尔实验室整体人员的质量有所影响。我想 IBM 在那个时候也在做同样的事情。还有一些公司（如甲骨文、Sun、惠普等），他们从来没有意识到基础研究可以培养很多人才，他们仅仅花两年的时间做基础研究，38 年的时间来做一些对公司有价值的事情。

玛丽安·温丝特：您最引以为荣的研究工作是什么？

杰弗里·乌尔曼：（稍微犹豫了一下）双向确定性自动机。

玛丽安·温丝特：为什么这个工作最令您满意呢？

杰弗里·乌尔曼：它是最酷的定理。Aho 和我一起花了三天的时间从各个角度进行研究，最终得到了所有的东西。奇怪的是，激发我们对这一理论进行思考的竟然是吉姆·格雷（Jim Gray）。他的毕业论文是关于双向确定性下推自动机。下推自动机在上下文无关的语言中是很常见的，但是在输入端你可以进行双向操作。吉姆在对这类语言的性质进行证明的时候，利用了大量的下推自动机的理论。我们不知道他到底是怎么做的，我们仅仅知道结果，

并且设想在更一般的情况下如何实现，而不是仅仅利用下推式自动机特有的一些性质。如果你问 Aho 的话，他会告诉你这些的。

玛丽安·温丝特：这也是他最得意的工作吗？

杰弗里·乌尔曼：是的。

玛丽安·温丝特：您作为一名计算机科学研究人员，如果您可以改变自己做过的一件事，您希望改变什么？

杰弗里·乌尔曼：我认为我最大的失败就是做了太多纸上谈兵的事情，而没有去做一些实事。现在我年纪太大了、身体胖了、也懒惰了，所以也做不了什么事情了。其实在早些时候我应该严格要求自己去做一些事情的。

玛丽安·温丝特：您认为现在最令人兴奋的技术趋势是什么？

杰弗里·乌尔曼：最令人激动的趋势？

玛丽安·温丝特：哦，也不一定非得是最令人激动的。

杰弗里·乌尔曼：有很多事情都在发展之中。最重要的一件事情可能是正在建立的互联性（Connectivity），技术的发展最终将使每一个人都成为世界网络的一部分。我认为很多技术趋势对于真正实现网络世界非常重要，而我们在实现这一目标的过程中刚刚走出低谷。例如我昨天刚刚在报纸上读了一篇文章，文章介绍了 98% 的年轻人如何使用短消息，并且 13% 的人使用短消息进行约会。我们的生活以及我们彼此联系的方式都在发生巨大的改变，并且变化很快。这不仅仅是数据库的问题，它是计算机、通信、网络和数据库的一个综合性问题，无线技术将对我们彼此之间的联系方式带来巨大改变。

玛丽安·温丝特：那是不是意味着数据库研究人员有些该做却没有做的事情啊？

杰弗里·乌尔曼：不一定。很明显，数据库的界限一直在演变之中：如什么样的人算是数据库研究人员，什么才是数据问题。我在很多场合多次谈过什么是数据库问题，什么是数据库技术的核心，我们必须拓宽视野。我也不会对什么是数据库问题什么不是冒然作出评价。

玛丽安·温丝特：我不是那个意思。我的意思是有些确实需要有人进行关注研究的问题，实际上没有人在研究。

杰弗里·乌尔曼：有人在关注这些领域。例如，很多人仅仅是从某一个方面关注无线通信。我不知道这个问题在数据库领域是否严重。我看到有些数据库方面的工作是关于局部性维护、个人信息追踪。这里面有很多有趣的研究问题。我认为没有必要所有的人都去研究这类问题，或者是宽带通信问题。有一些人在研究流媒体数据库，主要解决视频流的传输速度问题。这些都是很有趣的问题，但是没有必要所有的人都去研究这些问题。

玛丽安·温丝特：您对那些初学者、研究人员以及专业人士有什么建议吗？

杰弗里·乌尔曼：可以。

年青人开始时必须得知道自己到底在做什么。你的目标是什么？是为了名誉、金钱还是快乐？过去我经常玩一种棋牌游戏，玩的时候你必须确定自己的目标是什么，目标可能包括名誉、金钱和快乐。这是人生中最重要的大目标。你必须事先决定哪个对你来说是重要的。一旦确定了自己追求的目标，就要确保付诸行动。要有闯劲，勇于挑战。

玛丽安·温丝特：非常好，谢谢！

（马友忠 译，孟小峰 审校）