

瑟格·阿比特博访谈录

本专访主要介绍了：如何在欧洲着手建立一个研究团队，为什么系统文章不必包含性能测试部分，对象数据库的价值等。

玛丽安·温斯特

欢迎来到本期SIGMOD Record数据库界杰出成员系列访谈。我是玛丽安·温斯特，今天我们在芝加哥伊利诺斯州的SIGMOD2006大会上。我们邀请了法国国家信息与自动化研究所的资深研究员，Gemo数据库团队的带头人瑟格·阿比特博教授。他主要专注于数据库、Web数据和数据库理论方面的研究，曾经荣获1998年SIGMOD创新奖，并是提供基于XML内容管理的Xyleme公司创始人之一。他在南加利福尼亚大学获得博士学位。欢迎瑟格！

玛丽安：瑟格，你已经建立了一支欧洲最成功的数据库研究团队。你是如何做到这一点的？与在美国相比，在欧洲建立一个研究团队所面临的挑战是否不同？

瑟格·阿比特博：这个研究团队的成功是许多人的功劳。该团队的前身是Francois Bancilhon和Michel Scholl赞助的Verso研究团队。因此，当我到来的时候，这支团队已经存在了。我所做的只是引进了一批新的、充满活力的、非常好的科学家——Luc Segoufin, Ioana Manolescu, Sophie Cluet。最近，我们搬到法国国家信息与自动化研究所的一个新研究单位。我的想法是研究团队应该尽量和大学保持密切的联系，所以现在我们的团队在厄尔萨伊大学，并且和Christina Rousset领导的知识表示研究团队合并。除了需要接近大学外，我们还将数据库系统和知识表示研究领域的专家和学者集合在一起，因为我认为这是真正解决Web问题的需要。

玛丽安：根据你的经验，对于那些试图在欧洲组建研究团队的人，你有什么建议？

瑟格·阿比特博：这很简单。与在美国建立一个大的研究团队一样。你必须将一些人才集合到一起，找到那些最好的研究人员并说服他们加入你的团队。这并不容易，但这就是你所要做的。

玛丽安：欧洲的数据库研究和美国有何不同？作为一个欧洲的数据库研究者，你有没有因为没生活在北美而有被冷落的感觉？

瑟格·阿比特博：数据库研究在欧洲起步比较晚。我在美国攻读了博士学位，当我回到欧洲时，这里数据库的研究团队还非常少。总的来说，在美国这个有趣的研究工作仍然在继续，我们必须尽快赶上。在某种程度上我认为欧洲现在已经迎头赶上了。

另一个区别是主要的数据库企业都在美国，所以欧洲的工业界要有所作为是非常不容易的。另一方面，既然我们已经在欧洲建立了一个很强的数据库社区，没有大的数据库公司对我们来说也许是一个优势。我相信未来许多有趣的研究工作将是Web数据方面的，这些研究可能是由一些刚刚起步的小公司来驱动，就像我们欧洲的那些小公司。

玛丽安：最近，我们已经看到美国几位著名的数据库研究人员回到了他们在欧洲自己的国家，如：Peter Buneman, Timos Sellis, and Yannis Ioannidis。这是一个趋势吗，如果是的话原因是什么？

瑟格·阿比特博：我认为这是一个很明显的趋势，而且从个人而言我很喜欢这个趋势。它的

原因是什么？欧洲的数据库研究目前正处于一个非常合理的水平，其经费支持也变得越来越好。美国的政治形势也许可以在某种程度上解释这一点。目前在美国，你们的政府不能吸引太多的研究者了。

玛丽安：你说是因为伊拉克战争，还是由于对数据库研究降低了的资金投入，或者是两者都有？

瑟格·阿比特博：我认为都有，但是主要是政治原因。

玛丽安：你的研究团队是世界为数不多的研究成员大部分都是女性的研究团队之一。为什么会这样？

瑟格·阿比特博：原因之一是Francois Bancilhon从来没有在男性或女性研究人员之间做出区分，他总是想要获得最好的研究人才。他为团队找到了非常好的女性研究人员，并且我也一直努力保持这个传统。自从成为这个团队的带头人，我就雇佣了四个常任研究员，他们中两男两女。我并没有故意而为之，只是选择认为合适相应工作的人员。也就是说，这不是我刻意的选择，但结果却是如此。

玛丽安：你的一个同事告诉我你是“女权主义者，但是一种法式的女权主义”。这是什么意思？

瑟格·阿比特博：谢谢！，我想我确实“非常女权主义”。我一直认为男人和女人应该获得同等的机会。在我的职业生涯中，有机会和Jennifer Widom、Sophie Cluet这样的女性一起工作，进一步增强了我的这一观点。那么，“法国式的女权主义者”，这是什么意思？也许与美国的女权主义区别是我们不试图相信男女都一样。我们认为男女有别，并且我们喜欢这种区别，但我们也鼓励男女之间的机会平等。

但仍有很长的路要走。例如，我最近参加了一个微软组织的讨论会——“Towards 20/20 Science”，其目的是建立计算机科学议程帮助广大的科学家。物理学家、化学家、生物学家——每一种学科的代表都有出席该讨论会。但我们发现参会者大多数都是男性研究者。这很荒唐，其实我们可以做的更好。有很多伟大的女性科学家，所以我们应该更多关注男女平等问题。

玛丽安：你是否建议在美国的计算机科学博士毕业生考虑在欧洲工作？

瑟格·阿比特博：当然！我认为在欧洲获得一个好的大学的职位要比在美国更容易。我想去欧洲是新毕业的学生应该考虑的事情。而且，欧洲也很好，应该会给他们一个很好的经历。

玛丽安：最初你是一位数据库理论家，但目前已越来越转向实践方面的研究。最近，你甚至加入一个新成立的公司。这是怎么回事？我更愿意接受你理论家和实践家的双重身份。

瑟格·阿比特博：这个故事很长，涉及到很多方面。我就讲其中的一部分：它是如何开始的？当时，我在斯坦福作为为期两年的访问。我的朋友Francois Bancilhon对我在斯坦福的研究工作非常感兴趣。由于他是工业界的，所以当我向他解释半结构化数据时他问我：“这些工作对工业界有什么好处呢？”因此我们开始讨论这个问题，然后通过电话会议讨论创建公司的想法，我认为这是一个挑战。我相信半结构化数据非常有用，我必须证明这一点，至少应该向Francois Bancilhon证明。

后来我们邀请Sophie Cluet也加入其中。我们研究这个课题一年，并开发了一个软件，就是后来的Xyleme系统。

启动项目和研究工作之间的相互影响关系非常有趣。项目的启动是基于我之前的理论研究，当我们创建公司时，我原以为研究工作就此结束，但结果并非如此。实际上Xyleme项目为我们带来了许多新的研究性问题，其中一些我目前仍在研究。

玛丽安：你发现的什么研究问题是你之前没有想到过的？

瑟格·阿比特博：以前，我们正在为Web搜索引擎开发一个页面排名算法。这个问题非常有趣，但它需要的很多资源我们没有。例如，Google可以使用很多机器存储Web链接图，但是我们没有。我认为肯定有一种不需要很多资源就可以完成排名工作的方法。我们和一些学生一起开发了一个在线算法，该算法不需要存储Web链接图就可以计算网页的排名。然后，我们进行了算法分析，并开始研究当Web链接图演化时出现的问题。所以还有很多开放性的问题。

玛丽安：当你描述你的公司时，我没有看出它与基于XML的内容管理之间的联系。那么，这里什么与XML相关？

瑟格·阿比特博：要明白这个问题，我们必须回到那个时候。当XML刚刚起步时，我们有了一个疯狂的想法：XML可以征服这个世界。我们认为五年以后，每个人都可以在Web上发布XML，我们打算为Web上的所有XML提供一个查询引擎。因此我们的目标是能够发现、索引和查询数亿个XML页面。

当然我们是错误的。但后来我们发现，虽然Web上没有如此多的XML，但在许多公司里他们有很多XML。所以我们改变了商业模式，现在我们有了一个产品，它可以发现、索引和查询公司里的所有XML并且可以使用语义标签、语言分析等来丰富其内容。该产品的可扩展性很好，因为我们最初的目标是针对Web上的所有XML，但公司通常并没有如此多的XML数据，所以这个产品即使有非常大的数据量也会运行得非常快。

玛丽安：你的页面排名算法在某种程度上引人注目吗？

瑟格·阿比特博：页面排名算法已经被放弃了，它仅仅是一个很好的研究性问题。

玛丽安：你为什么没永久转到工业界？

瑟格·阿比特博：在Xyleme这样的小公司里，从工程的角度看，开始是非常有意思的。因为你正在做一个产品，正在做一个系统，这非常有趣。你要去见客户，这也很好。但是过了一段时间，这就变得很枯燥了。你有很多可以改进产品的好想法，但管理者可能告诉你，这太昂贵以至于不能做这些改进工作，更糟是顾客并不需要这些改进。但如果客户并不知道这些问题，他们怎么想要这些改进呢？如果你不做任何新的东西，一段时间后这就会变得非常枯燥。

而且，从一个顾客的角度出发，你总是试图从事重复的销售，卖相同的东西，这与研究工作背道而驰。在研究中，一旦你已经完成了一些研究工作，你就不想再重新做一遍。所以我的经验是在一个新的或者小的公司里，有趣的部分是市场，工程部分很快会变得很枯燥。

玛丽安：你看到数据库理论界和系统界之间的不同吗？例如，功能评价或性能评价的方式有何不同？

瑟格·阿比特博：是的，我认为有很大的不同。从某种程度上说，一个理论的评价非常容易。你只需要检查定义是否严谨，证明是否深入，据此就可以评价这个理论。

当开始研究一个系统的时候，我认为将会学习一个非常有趣的新文化。但从某种程度上令我失望的是系统研究的评价方法。系统的评价比理论的评价困难得多。人们需要提供系统性能的评估，但我觉得当你认真检查实验时，大部分实验结果都不重要：你所发现的正是你所预期的，不同方法的比较是非常困难的。我认为，系统评审中有很多的随机性，其远远多于理论评审，而且我没有看到有任何方法可以得到改善。

系统研究方面的会议总是有一条规则：如果你的文章中没有性能评估，则将不可能发表。我认为这很愚蠢，因为多数情况下你看到的性能评估实际上是一些模糊的实验，这些实验室是由几个学生在几个月内突击完成，并不没有什么实际的意义。我认为性能评估是非常重要的，但真正的评估需要更长的时间和更多的博士生来完成。所以，我宁愿看到一些针对思想和功能评估的系统论文，而将性能评估问题留给那些真正讨论优化问题和性能问题的那些论文。

玛丽安：我认为所有好的系统思想都很肤浅。相反，一个高深或复杂的思想在实现的时候往往行不通。这种矛盾也可能会影响评估过程。

瑟格·阿比特博：你认为Google的PageRank思想很肤浅吗？

玛丽安：当然，它是一个伟大但又简单和肤浅的思想。如果你不能用两句话来描述一个思想，你就不可能真正的构建出这个系统。

瑟格·阿比特博：我对PageRank思想非常尊重，因为每个人都可能想到它。

玛丽安：是的，它是最好的系统思想的典型！它很浅显，每个人都可能想到它。现在回想起来，它看起来如此显而易见，但没有人去做。

瑟格·阿比特博：但是你必须首先想到这个方法，而且必须相信该方法有效，并且必须使其有效。这就是一个好的系统思想。它必须简单，但你必须证明它有效。我认为这需要工程学、好的思想和信念。

玛丽安：但在系统论文中，你如果不去实现想法，如何判断它是否有效？如果你实现它了，就可以将其作为判断想法是否有效的一种评估方法。

瑟格·阿比特博：我更倾向于使用原型系统进行评估：你可以做一个系统来说明思想可以实现、有合理满意的性能，且一切功能都具有，确保没有遗漏任何重要方面。除了要求有好的思想并将其实现外，要求具有性能评估是非常荒唐的，因为评估原型系统需要大量时间导致没有时间花在功能的添加上。我更感兴趣功能和可行性的证明，最后才会注意性能问题。当然，如果你所研究的问题是XML查询优化，则没有性能评估部分的论文是没有任何意义。但如果你有一个新颖的思想，我认为可行性证明就已经足够了。

玛丽安：有时做测试是因为读者将想知道他们将必须为你的新功能付出什么代价。例如，如

果采用你的新技术来替代旧方法，则他们可能必须放弃10%的性能。通过表明不必为了获得一个新的功能而放弃太多的性能，你从而可以显示你的想法有多么好。

瑟格·阿比特博：当然，事情有时候确实这样，但是提供测试并不应该是一个规则。

玛丽安：你已经对一些当时冷门的或有争议的领域做了很多的研究工作，例如嵌套关系、对象数据库和半结构化数据。你如何选择新的研究课题？

瑟格·阿比特博：我总是喜欢研究一些新的东西。我非常喜欢一些新的研究课题。可能这是因为我很懒：如果你进入一个新的研究课题，你没有无数的文章要读。当然，如果大家更愿意研究新的课题话，这将会是一个噩梦，那么我将不得不选择一个不同的方法。

最终，我会根据我想要一起工作的人来选择一个研究课题。我选择一个能给我带来乐趣的研究课题。所以乐趣和快乐是首要标准。

玛丽安：Mike Stonebarker将对象数据库称之为“一个零亿美元的市场”。这是否意味着，研究界不应该去研究它们？

瑟格·阿比特博：这两句话之间有一个很大的矛盾。Mike Stonebraker比我更了解工业界，这句话是针对产业界的，它和科学界无关。Mike关于零亿美元的声明绝不是针对对象数据库研究的科学价值。

实际上，我甚至不同意对象数据库在市场上是一个彻底的失败。我的妻子Sophie Gamerman是O2 Technology公司的副总裁，我们家的部分收入来自O2 Technology公司。所以，我真的很感谢对象数据库产业！

现在，从研究的角度来说，我认为对象数据库已经带来了许多非常好的想法，它们已经给数据库领域带来了很大的影响。例如，在XML中对持久化XML我们经常使用文档对象模型接口（DOM）。在我看来，DOM就是一个对象库。所以当你正在做持久化DOM时，无论你喜欢还是不喜欢，你正在做的就是对象数据库。

有趣的是，你不应该说你正在做对象数据库。一些风险投资家邀请我为一个新成立的公司做一个技术尽职检查。在听了该公司创始人的半个小时的陈述后，我告诉他们，他们正在做的已经被Object Data Management Group(ODMG)实现了，并问他们意识到这一点。他们告诉我，他们知道他们正在做对象数据库，但是他们不愿意提及，因为这不是筹集资金的一个好方式。这些人正在做对象数据库，他们知道对象数据库技术，但是他们不想提对象数据库，因为一直以来人们都认为对象数据库是一个不好的技术。对象数据库不是一个成功的产业，但是他们是一个非常成功的技术。

玛丽安：在法国任终身教席如何？

瑟格·阿比特博：这与美国非常不同。一旦你完成了博士学位，你通常必须做一到两年的博士后。然后你会在大学或者像INRIA这样的研究机构中获得一份永久性的工作。但是你并没有真正的一个终身教席系统。

玛丽安：那好吗？

瑟格·阿比特博：我不认为这是好的。我认为这有点太早了，以至于不能看清楚一个人是否真正喜欢研究，是否真的擅长研究工作。在美国，对争取它们的人来说在终身教席之前这段时间可能是一个巨大的压力。但另一方面，五到六年后当某人获得终身教席时，你知道他是个专业的研究人员。

玛丽安：有人建议我问你是否认为Xquery很糟糕。你愿意对此作出评论吗？

瑟格·阿比特博：我知道现在指责Xquery和XML模式非常流行——在SIGMOD06上我也听到了一些这样的指责。但我不想参与其中。

如果是我设计Xquery，我想会有一些不同。我会让它更加实用一些，让它进一步远离SQL。但我事实上我并不是研究领域中的一员。该领域的研究者们一起提出了一个方案，这是一种妥协，所以它并不完美。但是至少我们有了标准，有标准比没有好。

我认为某种程度上，Xquery的主要问题是忽略了实际问题。Xquery 可以用来查询一个本地XML数据库，这并不是一个实际问题。XML最初是作为Web的数据交换语言被提出来的，所以我们需要的是种语言，它可以用来描述分布的XML资源和一般的分布数据资源，并且可以查询它们。过去几年我和一些同事Omar Benjelloun, Ioana Manolescu, Tova Milo等致力于这方面的研究。一些人研究XML存储、XML处理这是很好的，但是我认为在Web内容的分布式查询处理方面需要有更多的研究。

玛丽安：研究者的性格和人格对他们的成功有何影响？

瑟格·阿比特博：研究主要涉及到与人打交道。虽然有些人是独立研究，但大多数人属于某个研究团队。你的个人素质会影响整个研究团队。一个团队的研究成果应该大于每个单独成员的研究成果之和，要做到这一点需要的不只是智力方面的人才。它还需要一些人才，他们可以其他人解释，倾听他人在说什么，努力和别人一起工作。这并不容易。就个人而言，我有不容易合作的名声，但是另一方面，我和许多合作者成为非常好的朋友。

玛丽安：你最近出版了一部小说《Sparrows on the Web》。这本书中的计算机科学家何种程度上是受到现实生活的人或事的启发？

瑟格·阿比特博：书中的一部分是描述一个开发搜索引擎的创业公司，现实生活中我接触过这样的一些人。当然，我书中的人物往往受到我认识的一些人的影响。但是你不必想得太多，不要试图认出任何人。有时候人们试图在认出我书中的一些角色。曾经我收到一位女士的邮件，她认为自己是书中的角色之一，此前我从没有见过她。所以不要在我尝试在我书中认出某人。

玛丽安：除了写数和小说，你还爱好雕塑，对政治有浓厚的兴趣，并且还很顾家。你是如何安排时间做到这一切的？

瑟格·阿比特博：我是个超人。你不要告诉任何人呢，但这就是我。

玛丽安：对那些初出茅庐或者处于职业生涯中期的数据库研究人员或从业人员，你有什么建议吗？

瑟格·阿比特博：如果你不认为你是一个富有创造力的研究者，那么你最好去做一些其它的

工作。你应该去做一些更容易的事情，例如：开发系统，或者管理工作。

玛丽安：在你过去所有的研究中，有没有你喜欢的工作？它也是最有趣的研究工作吗？

瑟格·阿比特博：是的，确实有一个工作我很喜欢，那是不久之前和Victor Vianu合作的工作。那时我们正在研究定点逻辑，令我们真正感到困惑的是有一些不能使用一阶逻辑表示的简单查询。我们一直对此课题研究了一段时间，并且写了多篇论文。当我们设计了等价类的概念后，一切问题突然迎刃而解。之后我们定义了一些定理，这些定理就像定理“如果定点逻辑与偏定点逻辑相同则P-TIME等于P-SPACE”。没有人关心这些定理，但是对我们来说其伟大之处就是对这些等价类的理解。当我们理解它的时候，我们认为它是美丽的。我真的很开心，我想Victor也共享了那段美好的时光。

玛丽安：如果你给你足够额外的时间做其它一些目前你没有做的研究，你将会做什么？

瑟格·阿比特博：我不需要额外的时间。如果有额外时间的话，我将会写更多的论文、读更多文章、做更多管理性的事务。所以如果不得不选择的话，我宁可少一些时间。

玛丽安：如果你能选择的话，你希望让什么重现？

瑟格·阿比特博：我从来没有足够深入地做事情。写论文的时候，我喜欢写初稿，但不喜欢修改论文，我觉得那很枯燥但又不得不做。我一直努力改善这一情况。

我已故的朋友Paris Kanellakis有一个教训，他与他的家人大约在十年前消失了。那时我和Paris正在一起研究IQL，它是对象数据库的一个形式模型。他一遍遍的要求我仔细检查这个模型。我想我们在黑板上写了这个模型定义有40遍，但是每一遍只有一点点的改进。那时我有点恼火，因为我想要加快速度做进一步研究。但是，当我回过头来看它，我真的很喜欢这个工作，因为这个模型非常清晰，我们花的时间是值得的。因此，我要做的改变是将工作做得更加深入和彻底。

玛丽安：非常感谢您今天的谈话。

瑟格·阿比特博：谢谢您的邀请。

（马如霞译， 史英杰校）