

第17篇

---

# 韩家炜访谈录

*Interview with Jiawei Han*





## 韩家炜简介

韩家炜 (Jiawei Han), 伊利诺伊大学厄巴纳 - 尚佩恩分校计算机科学系教授。1978 年考入中科院研究生院, 1979 年考取首批公费留学赴美, 1985 年于威斯康星大学麦迪逊分校获得博士学位。韩家炜教授主要专注于数据挖掘、信息网络和数据库等方面的研究, 先后在 KDD、VLDB、SIGMOD 等发表高水平论文 400 余篇, 并在 2002 年 ICDE 会议上被授予最杰出贡献奖, 获得 2004 年 ACM SIGKDD 最佳创新奖, 2005 年 IEEE 计算机学会技术成就奖, 2009 年 IEEE 计算机学会 Wallace McDowell 奖。同时, 他创办 ACM Transactions on Knowledge Discovery from Data 期刊并担任主编, 他是 ACM 和 IEEE 会士, 美国信息网络研究中心主任。其与人合伙撰写的《数据挖掘: 概念与技术》为本领域公认的经典教材。他是数据库领域杰出华人学者之代表。

本专访介绍了韩家炜教授如何研究数据挖掘、  
隐私等问题，以及如何管理自己的学生。



玛丽安·温丝特

**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。现在，我正在伊利诺伊大学厄巴纳 - 尚佩恩分校，将要采访韩家炜（Jiawei Han）教授。家炜是该校计算机系教授，在加入伊利诺伊大学之前，先后在加拿大西蒙 - 弗雷泽大学和西北大学任教。家炜的主要研究方向是数据挖掘，他是 ACM Transactions on Knowledge Discovery from Data 期刊的主编，并写了一部非常畅销的数据挖掘著作。他从威斯康星大学麦迪逊分校获得博士学位。家炜，欢迎您。

**韩家炜：**谢谢，谢谢您，玛丽安。

**玛丽安·温丝特：**家炜，根据发表论文的记录，您的 H 指数（H-index）已经到达 76，也就是说您已经有 76 篇文章的引用次数超出了 76 次。从这一点上来说，虽然我的 H 指数只有 27，对此我还是很高兴的。如果我的 H 指数也能达到 76，我会更开心的。所以，对那些希望获得更多成就和希望自己的论文更具有影响力的人，您有什么样的经验要告诉他们？

**韩家炜：**实际上，这是我第一次听说自己的 H 指数达到 76。记得当时有人告诉我说一个 H 指数衡量指标什么的，我去查看了一下，发现当时我的 H 指数是在 54 ~ 55 之间。我从来不知道我的 H 指数已经达到 76 了。对于做

研究，我想我能说的就是，如果你能够选择一些令人兴奋的研究点是非常好的事情。这也许很艰难，也许没那么艰难。但我认为，选择一些有点新鲜感、有点意义以及有点令人激动的研究点并试图找出相应的解决思路去做，这样比较有意义。当自己写出一篇文章后，总希望有人阅读它，这也是自己继续创作的动力所在。所以，如果自己在写作过程中获得了快乐与感动，可能读者也可以获得同样的感受。当然，并不是每一篇文章都能达到这样的效果。对于我来说，我很可能喜欢自己写的每一篇文章，但我还是偏爱那些比较有意义的具有吸引力的文章。由于我还指导很多学生，你不能期望每个学生都能写出受人欢迎的文章，这是非常困难的。也许我们会写出一些意义不大、缺乏兴趣的文章，但总体来说，我们尽量要从事一些有意义的研究主题，并试图找出相应简洁和优雅的解决方案出来。这样做读者会感到愉悦并引用你的文章。

**玛丽安·温丝特：**接下来让我们谈一下数据挖掘。数据挖掘在今天是很重要也很流行的技术。但是它又是一个非常年轻的领域，数据挖掘的第一篇文章大概出现在 1992 年的 KDD 会议上。是什么让您与其他 KDD 会议发起者开始从事这个新领域的研究工作？

**韩家炜：**你刚才说的第一篇 KDD 文章，大概是在 1992 年的一个数据库会议（例如，SIGMOD、VLDB 会议）上出现的。但事实上，数据挖掘的文章可以追溯到 1989 年由夏皮罗（Piatetsky-Shapiro）教授组织的第一届知识发现研讨会上。所以我还记得这事，是因为我投了一篇稿子过去，并参加了那次研讨会。当时的参会人员很少，大概只有 30 人。实际上，那次研讨会是附属当年在底特律召开的 IJCAI 会议的，当时每个参会的研究者都感觉数据挖掘是条“大鱼”，一个大的研究方向。我自己也深有体会，因为我的博士论文就立足于数据挖掘。实际上我当时从事的研究是演绎数据库。在当时，逻辑程序设计和数据库是非常热门的方向。

兰迪·凯兹（Randy Katz）教授对我从事数据挖掘研究有着非常大的影响。当时兰迪是威斯康星大学的教授。记得在 1983 年或是 1982 年，具体哪一年我忘记了，他举办了一次数据挖掘研讨班，当时日本制定了第五代计算机工程项目。他在研讨班的讲桌上放了一把日本军刀，并说这是日本人的挑战。他说日本人已经打算开发出一台能够支持高并行数据库搜索和推理的机器（Prolog machine）。他还说我们必须应对这样的技术挑战。我对当时的情景印象非常深刻。从那时起，我确定了自己的研究方向。

**玛丽安·温丝特：**对于数据挖掘来说，围绕它的应用领域是什么？

**韩家炜：**应该说数据挖掘技术可以应用到任何地方，你也能看到该技术的存在。例如在网页搜索过程中所使用的 PageRank 或 HITS 算法。本质上，PageRank 和 HITS 算法就是在做数据挖掘，因为如果有一个网页被其它网页指向，那么该网页就携带有语义并且具有重要的意义。这就解释了为什么最终能够找到我们感兴趣的相关网页。所以，我们首先应该认为数据挖掘是隐式的技术，该技术可能是最有趣但不一定是最有效的挖掘方法。但是，很多人都使用它。你想象一下亚马逊网站，他们会告诉你“人们买这本书的同时也会买其它书”。该网站就是使用了类似于协同过滤的算法或者一些其它的数据挖掘算法来达到这样的效果。你还可以想象一下 Google，人们搜索网页的时候，他们会用到一些有用的挖掘结果。所以对于这些不可见的数据挖掘技术，即使这些网络公司说他们不是真正地做数据挖掘的，但是实际上他们在使用类似的数据挖掘方法。我认为这可能是我看到的最有意思的应用领域。将会有越来越多的应用领域出现。

**玛丽安·温丝特：**您认为会有那么一天，数据挖掘方法不再依赖领域，就像一个统一判别排序模型，它不依赖于语义信息，不依赖于实际应用吗？例如，如果您看到一个客户的行为很不正常，那这个客户可能是在欺诈，或许也有可能他是一个挥霍浪费的人，又或许是一个噪音数据。您是否看到过

不依赖领域的针对基本数据挖掘任务的方法？

**韩家炜：**实际上，数据挖掘就像规则一样，你想找出一些通用的规则。某种程度上来说，你不想做得很具体，你知道许多方法都只能解决一些特殊问题。你想找到一个通用一点的不依赖于特殊领域的方法。但是不同的数据种类，需要不同的挖掘方法。例如，序列挖掘中所包含的数据类型：事务数据库中的序列，像购物序列；生物数据中的序列，像 DNA 序列和生物序列；或者文本序列。这些序列都有着不同的特点和模式。所以，如果你的算法能够针对这些序列数据挖掘出所有的序列模式，可能该算法什么也做不了。你知道它根本不具备通用性，更不能挖掘出所有的模式。在某种程度上来说，我们可以先说算法是为某种特殊应用而特制定做的。然后，你可以提出一个更有效的算法，扩展到其它应用范围。此时，你可以认为你所设计的算法对领域的依赖减小了。

我记得，我们提出过一个方法，好像是叫做 CloSpan，就是这个。在西蒙－弗雷泽大学时，裴健和我首先提出了 PrefixSpan 算法，紧接着严锡峰又提出了 CloSpan 算法。我记得普度大学的一个教授采用过这个算法，实际上他只是使用它解决生物数据的问题，并且发现了一些有趣的序列模式。我记得在我们数据挖掘研讨班的前几周，有几个日本研究者也在做着类似的研究，他们更多关注网页日志或者网页博客的挖掘。他们首先使用的是 PrefixSpan 算法。我并没有真正实际地了解，他们最先使用的是这个算法，某种程度上说明这个算法也可以被用在其它领域。所以我认为该算法变得越来越不依赖于领域。但是，我认为对于特殊问题，我们不得不先针对此提出一个专用方法，然后再想办法把所设计的算法不断泛化到其它领域。

**玛丽安·温丝特：**那就是说，数据挖掘是否像数据库那样也有一些通用的准则呢？

**韩家炜：**我认为，在数据挖掘领域，这个问题与数据库领域尤其是与关

系数据库领域有所不同。在关系数据库领域中，数据实际上是结构化的。对于这些结构化的数据，你可以很容易地做选择、连接、查询处理或者事务管理等操作。实际上，如果我们仅仅处理那些高度结构化的数据，那么你可以很容易地将一些现有算法移植到不同的领域中。但即使这样，人们也在寻找不同的模式。对于同样结构化的数据，无论是你想要发现类簇，还是做回归和演化分析，你可能会发现一些不同。由于人们发现的知识不同，所以这些算法很可能将被调整，或者对一些不同应用来说，这些算法可能有很大不同。这就是为什么一些人梦想着我们能够开发出很多通用的数据挖掘工具。你只需要下载它，并且希望它对任何模式都适用。至少，在这一点上，我认为是不现实的。所以，你必须更好地了解这个领域，必须真正知道：你想要发现的模式是什么，你使用知识的方法是什么。我认为，这并不像关系数据库那样仅仅使用简单的语言（如 SQL 或 SQL 挖掘）就可以解决所有的问题。

**玛丽安·温丝特：**好的！有两个人建议我问您关于数据挖掘道德伦理方面的问题。其中一个人给了我一个 ChoicePoint 的例子。我在维基百科上查看了 ChoicePoint 的相关介绍，上面说：“ChoicePoint 是一个数据收集公司，它为政府和工业界提供隐私情报服务。ChoicePoint 从多个公开和非公开的数据库收集个人信息卖给政府或者私人机构。这个公司拥有超过 17 亿的个人和商业数据，据估计这些数据被卖给 10 万个客户。然而，这些数据并不足够安全。该公司已经因拥有不准确的数据而面临一些法律诉讼，接受调查是否允许政治偏见影响政府合同的执行等问题，并且被指控非法将外国公民的数据出售给美国政府。”当然，在挖掘的结果中，总是存在一些错误和不准确的信息，并且总是存在一些贪心或易受贿赂的人想利用这样的挖掘结果。对于数据挖掘界来说，我们应该如何解决这样的问题？

**韩家炜：**实际上，在数据挖掘领域中，我看到大量新闻和一些有争议的事情。所以我要说的第一件事情就是：对任何研究来说，例如当你申请一个自然基金项目的时候，他们可能问：“你处理的对象是人类还是非人类？”，

数据挖掘实际上处理的是这两种类型对象。很多数据挖掘问题处理的并不是人类对象。例如，如果你试图挖掘一些天文图片（像吉姆·格雷所做的，他参与建设的是天文学数据库），你仍然需要大量的数据挖掘。你可能从不担心暴露任何星星的隐私。所以从这方面看来，没有任何隐私问题。实际上，在世界的任何地方数据都可以是公开的，任何人都可以共享。所以，有很多不需要担心隐私问题的数据挖掘任务。

然而另一方面，数据挖掘中有一些人类研究对象。例如，你挖掘一些和人类有关的数据。一旦我们开展这方面的研究，就必须考虑到数据隐私和安全这一类的问题。我个人感觉在进行数据挖掘时，我们通常可以有两种思路。一种思路是使用内部的数据挖掘软件自行对数据进行分析。例如，沃尔玛有一些数据挖掘人员，他们坐在沃尔玛的数据库办公室进行数据挖掘。同时，他们需要选择哪些数据可以发布，哪些数据不可以发布。然而在内部数据挖掘中，也存在你是否能看到个人数据的问题。一旦你设定了内部数据挖掘的一些合适的标准，例如，你可以说，我挖掘一个客户的记录。甚至你可以利用一个信用卡号的消费记录来挖掘客户的消费行为。你会说，参与数据挖掘的人员被禁止利用该信用卡号进行其他的数据关联，卡号也不会被泄露。我个人认为，这种方式没有泄露任何人的隐私。而关键的问题在于，你到底能够发布什么内容？例如，你可以采用一种统计的模式来发布数据，就像美国统计局那样。他们定期发布很多数据。你可以买 CD，然后可以拿到多年的数据。你可以查找邮政编码，但邮政编码数据量也相当大。一般来说，一个邮政编码可能会涉及到成千上万的人。如果你采用 k-匿名，k 的值可能会达到上万。因此你无法从这些数据中建立联系，因此发布这类数据是安全的。

我的观点是，如果你进行内部数据挖掘，并且谨慎地发布数据，保持你的 k 值或者其他隐私保护参数足够大，数据仍然是安全的。你可以使用这些数据。我想很多人都在使用这些数据，就像美国统计局，他们发布大量数据，很多人正在使用它们。我想这些数据是非常必要的，因为管理者（例如奥巴马）想知道这些统计数据。任何做决策的人都需要这些数据。为了保证这类

数据的隐私，你保证  $k$  足够大就不必担心了。

**玛丽安·温丝特：**可是  $k$  值越小，那么你的产品收费就会越高啊。

**韩家炜：**是的。但是那样会有危险。如果  $k$  值太小了，人们可以识别出一些敏感信息，进而威胁人们的隐私安全。这也是为什么数据挖掘和数据发布中的隐私保护如此重要的原因。人们想挖掘深入的信息，而同时又要保护人们的隐私，这是两个矛盾的目标。另外一个很重要的事情是开源数据挖掘，我认为开源数据挖掘会更加危险。

**玛丽安·温丝特：**什么是开源数据挖掘？

**韩家炜：**开源数据挖掘意味着你上传数据，然后其他人来进行挖掘，而你又不想让进行数据挖掘的人挖掘出超出你需求的信息。我认为这种事情很容易失去控制。这个方面有大量的研究性论文，一些论文说可以实现  $k$ -匿名、 $l$ -多样性、 $t$ -近邻性，或者实现所有的这些算法。而来自于得克萨斯大学奥斯丁分校的一个教授表明：如果你可以保证这些都是安全的，然后你做数据挖掘，你很可能找不出真正的模式。这样有可能导致挖掘者挖掘出的信息还比不上一般的入侵者得到的知识信息。我认为这是很有道理的。如果你对数据进行处理，使其没有任何特征，然后交给其他人进行挖掘，可能找不到任何有意义的模式。但是另一方面，如果你直接把一些粗糙的数据或包含敏感信息的数据交给第三方进行挖掘，这也不是一个好想法。

**玛丽安·温丝特：**某些人评论说创业对于今天来说更加容易，那么我想知道当您计划创办一家公司，您想要做些什么产品。

**韩家炜：**首先，针对创业到底是容易还是困难的问题，不同的人可能会有不同的观点或者经历。我个人认为是这样：当然会有一些人在从事数据库或者数据挖掘的工作，这是一个非常实用的领域，会有大量的应用，而基于这些应用又会出现大量的创业公司。研究者是否需要去创业或者一些人利用

这些思想去开公司，不同的人有不同的观点。对于我来说，我确实喜欢做研究，原因是我对开公司不太感兴趣。同样，对于研究，我已经感觉到筋疲力尽了。如果我真的开公司，我可能就没有时间睡觉。

**玛丽安·温丝特：**在企业或者创业公司干扰的情况下，您认为带研究生有什么挑战吗？

**韩家炜：**对于一个学生来说，我认为在伊利诺伊大学厄巴纳 - 尚佩恩分校要比在大公司或者大城市好得多。比如，加州旧金山湾区中心或者西雅图的某些地方。这些地方非常容易吸引学生，因为他们只要打个电话，骑着自行车或者开车就可以到那里。我认为这里的学生并没有真正地被大公司分散注意力。某种程度上，对于一个学生来说，这是了解公司需求、学习一些东西的好机会，可以更好地了解一个真实的世界。我鼓励学生走出去，做一些暑期实习，特别是去一家真正的公司，去研究院做暑期实习生。这是非常好的实践机会，因为你可以学到一些关于应用、工业等外面世界的东西。当你回来的时候，你可能发现不同的研究问题，产生不同的想法，可以构建你的社会网络和研究网络。我认为这些能够很好地帮助学生。

**玛丽安·温丝特：**现在企业和学术的工作联系如此紧密，对于一些年轻的毕业生选择去学术界还是工业界工作，您有什么建议？

**韩家炜：**我认为不同的学生有不同的想法和喜好。一些学生喜欢去企业。我有一个叫邵铮的学生，非常聪明，可惜他没有完成博士学习。中间他首先受到雅虎公司的吸引，离开实验室去了那里，当然他做得非常好也非常成功。他几周前还回到我们实验室招聘新人。

大多数学生喜欢到工业界工作。但我希望他们能够拿到博士学位，因为一旦获得了更多的知识，掌握了一个领域的研究现状，还有博士学位，那么就更容易获得研究或者研发的职位，或者是能让人更好展示才能的职位。因此，我鼓励每个学生都能拿到博士学位再就业。

另一方面，我也知道有些学生是真的愿意做研究。我告诉他们关于做研究的两个选择：一个是去大学，因为在大学里做研究不仅仅是教书，你可以突飞猛进，和一群研究生一起工作，可以找到很多的研究兴趣点。另一个选择是去研究型实验室，比如 IBM 研究院、微软研究院、雅虎研究院、谷歌研究院，有很多这样的研究院或者研究实验室。我觉着这些研究中心也非常好，因为他们有很多具有博士学位的研究人员，他们非常有才能，而且人也非常好。和他们一起工作既可以真正接触工业界，也可以和那些到工业界进行合作研究的教授一起工作。这样就可以扩展自己的视野。比如，像严锡峰，他在 IBM 研究院做了两到三年的研究员，最近他以助理教授职位加入了加州大学圣芭芭拉分校。他在学校和研究院里面都有很好的经验。

**玛丽安·温斯特：**在中国的文化大革命期间，您只是一个年轻人，然后您马上就成为了威斯康星大学的计算机科学研究生，那么您是如何做到这种转变的？

**韩家炜：**这是一个很沉重的话题。我只能说，是这样的。在那个时期，这是一个不同寻常的转变。不仅仅是因为在文革时期我很年轻，历史在那时自觉不自觉地被严重扭曲。我的家人很明智，当时我几乎是处在社会的底层，我在农村劳动了好多年，很不容易。实际上，不仅是我。整个国家的大学关闭了 12 年之久，从 1966 年到 1978 年。对我自己来说，当然非常艰难。

转折的时机发生在 1978 年，中国恢复了研究生教育以及研究生院。我去中科院，并且通过了专业考试以及英语考试。直到 1979 年，中国刚和美国有了外交关系后，我就到了威斯康星大学读书。我是非常幸运的，但是过程很艰难。我只能说，中国关闭了 30 年的大门，那时候大学里能看到中国学生是第一次。我觉着威斯康星大学需要极大的勇气来接受这批中国学生，因为那时的中国没有 TOFEL 和 GRE，完全没有考试的体系。我记得有一个威斯康星大学的教授告诉我，他们之所以敢于接受这批学生，是因为他们看到中国曾经是一个大国，现在也是一个大国，那么肯定有很多有才华的人。

实际上，给我写推荐信的一位教授曾在伊利诺伊大学厄巴纳－尚佩恩分校（UIUC）获得了博士学位，非常不幸的是，他在文革中被打为右派，劳动改造，并禁止他做研究长达 25 年。然而，他仍然为我和陆宏钧写了推荐信。当我们最初来的时候，威斯康星大学实际上对我们这几个学生特殊看待，而不像其他正式的研究生一样，因为他们无法对我们做出判断。但是他们说，既然教授说我们几个很优秀，并且考入了中科院研究生院，那么我们就一定非常优秀。他们就这样盲目地信任了我们。进入学校以后，我们很顺利地度过了第一年，并且也都顺利毕业。

实际上，在那几年有很多类似的事情。我记得有一次周圆圆邀请她的导师李凯做一个报告，同时还邀请我和其他几位教授到她家做客。李凯见到我就开玩笑地说：“那几年能出来读书的中国人并不多，其中最多也就有四到五个成为 ACM 会士，而我们三个竟然是同学。”简单来讲，即使是只有我们三个人，竞争也是异常残酷的。我觉得，中国已经进行改革开放，国家本身有很多变化，同时给众多学生也带来了巨大改变。

**玛丽安·温丝特：**2007 ~ 2008 年本来应该是您休息的一年，而您却一直在继续做研究、写项目。为什么不出去走走或者做点别的事情呢？

**韩家炜：**我觉得有这么学生跟随我一起工作，我很难长时间离开他们。同时也有很多工作等着我去做。我觉得在那一年，我不仅做了很多研究工作，也对我的书做了一些改进，同时也做了很多其他事情。所以，我的时间比较有限，希望将来能有空闲时间好好休息一下。

**玛丽安·温丝特：**我有一个比较有趣的问题想问问您。

**韩家炜：**可以！

**玛丽安·温丝特：**下面这几件事情您觉得哪一个是最有意义的：（a）写一本有影响力的书；（b）培养出很多杰出的博士毕业生；（c）发明比较好的算法；（d）赢得很多奖项。

**韩家炜：**这有点难说，不过我觉得，培养出优秀的学生应该是最有意义的事情。因为不仅仅是你自己培养出了很多学生，同时这些学生也会成为整个学术领域的新生力量。如果你培养不出一批好学生，那么整个领域的问题也不可能由少数人全部解决。因此我觉得这是最有意义的。

**玛丽安·温丝特：**目前为止您有多少研究生、博士后以及访问学者？

**韩家炜：**这不大好计算，要看怎么来数了。我大概有 17 个博士、2 个硕士、1 个访问学者和 1 个访问学生。

**玛丽安·温丝特：**那么，刚才的回答对我的下个问题是一个很好的引导！有人告诉我，“我对于家炜的时间管理技巧和他对于研究的热情感到惊讶。即使面对他庞大的研究小组，他仍能够对绝大部分学生的电子邮件在数小时内做出回复，不管是在早上、中午还是午夜”。那么，在面对这么多邮件的情况下，您是如何做到的？

**韩家炜：**首先我认为对于学生的监督，面对很多的学生，有好处也有坏处。挑战在于当每个学生处于不同的情况时，该如何应对他们。从我的观点来看，你不需要自己去做每件事。学生能够协调工作。另一方面，学生自己也能够很好地组织起来。在很多情况下，无论是在加拿大还是在美国，我觉得总有一些学生，他们具有领导才能，他们将来极有可能成为领袖或是教授，因此他们能够真正地将事情组织起来。你一定要让这些学生扮演些角色。同时，我实际上是以小组的形式来面对很多学生的。在某种程度上，这些小组是动态的，你可以这样理解，一旦我们有些研究课题，有时候我会发个电子邮件问大家“谁对这些课题感兴趣？”。总会有一些志愿者，他们中的某些人是很热心的。我会让那些热心的学生作为领导，这样我们就形成了研究小组。我们在完成课题后，甚至在完成这个课题之前，会出现一些新的研究课题。一些精力充沛的学生能够加入 3 个或 4 个不同的研究小组。我认为这个可能会真正地减轻我的负担。

但是对于快速回答邮件的问题，我认为首先是我试着去回复，有时是试着去忘记，不然电子邮件会堆积起来，我不得不一遍又一遍地读它们。这也许不是个好的习惯，实际上我记得 Johannes Gehrke 在一篇 SIGMOD Record 的文章中写道，他尝试着不被电子邮件打扰。他试着将电子邮件汇集起来直到某个时间（比如 3 点钟）集中回复。我认为这是个好的习惯，因为你可以集中更多的精力在你的研究上。有时候我就被邮件打扰了。

**玛丽安·温丝特：**您有什么建议给新的或者处于职业生涯中期的数据库研究者及从业者？

韩家炜：是的！我认为一件可能的事是选择一个有希望的好的研究课题。通常，我会说职业生涯中期的研究者应该足够大胆去挑战一些新的事物。我当初选择数据挖掘在某种程度上是为了挑战自我。如果我觉得这是个好的课题，我就会投入。我认为对于职业中期的研究者，尤其是你已经获得终身教职了，你就应该有足够的勇气去寻找一些你觉得有挑战性的、令人激动的东西，然后全力投入。当然，与此同时，如果你原来在某些方面已经有很好的背景，你也需要继续进行下去。有时候你做一次完全的转换，但是有的时候你也许会转变回去，但是通过做一些新的东西，我认为总会给你带来更多的信心和能力。

**玛丽安·温丝特：**对于之前的研究工作，有您最喜欢的吗？

韩家炜：哦，当然，让你兴奋的东西不管别人是否喜欢它，你总会对它感兴趣的。比如说，早些时候我在做演绎数据库的研究，因为感兴趣我能解决像递归这样非常不规则的递归编译的问题。再比如，用非常规的方法解决 N-皇后问题，对于这个问题我也是相当兴奋。不过，当然，你知道，有许多新的研究课题可以研究。像我们最初研究的面向属性归纳问题。在当时我也非常兴奋的。后来还有一些研究课题，我相信我们通过这种兴趣积累的方式解决了频繁模式、序列模式、图形模式等问题，我认为这些问题都是令人兴奋的。而且这些问题的解决也印证了我们的做法。

**玛丽安·温丝特：**如果您奇迹般地有了足够的额外时间去做您之前没做过的别的一些事情，这些事会是什么呢？

**韩家炜：**你的意思是研究之外还是与研究相关的事情？

**玛丽安·温丝特：**都可以。

**韩家炜：**事实上，除了研究，我还很喜欢旅行、登山。在我年轻的时候对这些事我是相当热衷的。

**玛丽安·温丝特：**哦，我明白了。在您谈到与研究相关或不相关的事情的时候，我在想您指的与研究不相关的事情可能是兼做 SIGKDD 的大会主席或者其他事情。而您指的是完全脱离当前的研究工作去登山。

**韩家炜：**没错！但另一方面，在研究中，我认为你总是试图找到令人兴奋的东西去做。很多时候，我喜欢阅读，例如，阅读“科学美国人”杂志。我总觉得有很多很多不同的研究课题，你很想知道，你也可以尝试。比如，有许多研究工作可以从这些刊物上获得想法。你会觉得：“哦，为什么生物学家可以这样做，我为什么不能做同样的事情？”我想很多研究或知识是可以交叉的。

**玛丽安·温丝特：**作为一个计算机科学研究者，假如您能改变关于自己的一件事情，那将会是什么？

**韩家炜：**实际上，我年轻的时候，也就是文化大革命时期，我梦想成为一名物理学家，文革使得我没机会实现梦想。不过，我认为经历这些事也是很有趣的。尽管我没能成为物理学家，但是当我进入计算机领域，我真的喜欢上了计算机。

**玛丽安·温丝特：**太棒了，感谢您接受我们的采访。

**韩家炜：**非常感谢你的采访。

（张啸剑、马友忠、史英杰、霍峥、马如霞、王江涛、慈祥 等译，孟小峰 审校）