

知识工程——“大数据”环境下的企业基因工程

□北京亿维讯科技有限公司 谭培波

摘要

“大数据时代”是目前最为流行的词语之一，然而，大数据只是一种对现在信息爆炸时代的表象认识，并没有深入到信息时代发展的本质层面，北京亿维讯公司总工程师谭培波认为：知识才是“大”数据背后的“小”，而知识工程，本质上就是建立企业的基因工程，即构建企业“天生的”学习和适应能力。

知识是“大”数据背后的“小”“大数据时代”无疑是目前最为流行的词语之一，一大堆关于大数据的科普和专业的书如雨后春笋一般涌现出来，你无法拒绝它的诱惑，总想弄清楚它到底在说什么，总在担心大数据会把我们带到什么样的一个不可预知的未来。然而，大数据只描述了数据的规模大，是一种对现在信息爆炸时代的表象认识，并没有深入到信息时代发展的本质层面，因此，大家都在等待一个新的时代的出现。

这个时代就是在 DIKW (Data-Information-Knowledge-Wisdom) 概念中界定的 K 的时代，知识的时代。知识正是大数据之“大”背后的那个“小”。比如第谷测了一辈子的行星运动轨迹，记录的数据汗牛塞屋，后来他的学生开普勒又花了将近一辈子的时间整理出了三大定律，这个大数据终于变得小了一点，到了牛顿时代，一个简单的平方反比的万有引力定律，就把所有一屋的大数据的内容全说透了。

虽然知识的概念出现很早，但到底什么是知识呢？在科学上要清晰的定义非常难，现在也做不到。但其中有一种解释，我很赞同，即知识是信息之间的关联。关联代表着新物质和新性质的产生，而不是静态的两个物体机械地摆在一起。比如 H 原子和 O 原子关联在一起，产生的不是 2 个原子物理上挨得很近这么简单，而是产生了一种叫 H₂O 的新物质，它拥有它的母体、它的前辈所没有的液态流动的性质，因为有了这个物质，孕育了所有的生命，造就了地球这个生机盎然的大千世界。



古希腊先哲巴门尼德约公元前515年-前5世纪中叶以后，开创知识论的法门

在知识挖掘中常举的啤酒和尿布的例子，讲的就是通过关联获得知识的例子。“啤酒与尿布”的故事产生于 20 世纪 90 年代的美国沃尔玛超市中，沃尔玛的超市管理人员分析销售数据时发现了一个令人难于理解的现象：在某些特定的情况下，“啤酒”与“尿布”两件看上去毫无关系的商品会经常出现在同一个购物篮中，这种独特的销售现象引起了管理人员的注意，经过后续调查发现，这种现象出现在年轻的父亲身上。在美国有婴儿的家庭中，一般是母亲在家中照看婴儿，年轻的父亲前去超市购买尿布。父亲在购买尿布的同时，往往会顺便为自己购买啤酒，这样就会出现啤酒与尿布这两件看上去不相干的物品经常出现在同一个购物篮的现象。沃尔玛发现了这一独特的现象，开始在卖场尝试将啤酒与尿布摆放在相同的区域，让年轻的父亲可以同时找到这两件商品，并很快地完成购物，这就是“啤酒与尿布”故事的由来。

啤酒和尿布放在一起，代表了一种精巧的知识结构，是一种高级的人类智力活动的结果。发现关联就是发现知识，就是发现新物质、发现新性质，就是发现新的运营模式，就是发现人们未知的新世界，随着关联规模和深度的增加，也就扩大了人类认识世界和认识自己的能力，从而也就扩展了人类的生存能力和把握未来的能力。

大数据的根本目标就是要像挖掘万有引力定律一样，挖掘数据背后的规律，让数据由大变小，从而为我所用，创造出更多的经济、社会和科学价值。因此，我们认为知识时代才是

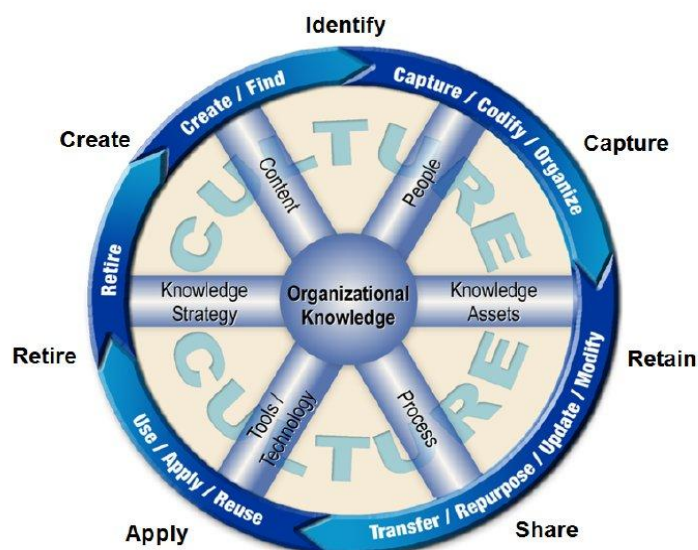
抓住时代发展本质的描述，而所有技术层面的新生事物，都是实现这个时代目标的物质和技术基础，比如搜索技术是信息时代的技术基础，网络是信息时代的物质基础。在新的知识时代，大数据和相应的挖掘技术是技术基础，而数据仓库和高速网络、智能终端等是物质基础。知识工程是实现知识挖掘的系统工程，是突破人类认识能力的局限，从“大数据时代”走向“知识时代”的基础之一。

知识工程，企业基因工程

所谓知识工程，本质上就是建立企业的基因工程，即构建企业“天生的”学习和适应能力。

首先，让我们先从自然界看看基因的伟大之处。2012年11月14日凌晨，8岁的“那仁”登上Nature杂志官方网站的首页头条新闻。这头来自蒙古国阿尔泰省自然保护区的雄性野骆驼，成为世界上首个向全球展示双峰驼全基因组序列图谱的野生骆驼。野骆驼是世界上惟一能靠喝盐水生存的动物。它能在极其严酷的环境中生存下来，正是由于其背后的遗传秘密。“野骆驼奔跑起来时速可达80公里，在遇到危险时，能连续奔跑两三天。它们机警而胆怯，视觉、听觉、嗅觉相当灵敏，顺风时可以嗅到数公里甚至几十公里外的气味，判断出很远地方的水源。它们可以在遇到沙暴之前，随时关闭鼻孔而不影响呼吸，来不及躲避就会平躺在地上避风。它们还会在风暴过后，用泪水冲洗沙尘以保护眼睛”。这就是这种“天生的适应能力”。

再来看另一个动物——狼。狼是生物系统中最有智慧的动物之一。CCTV曾报道，四川一位画家收养了一只小狼，养到3个月大的时候要放归草原，必须锻炼它捕杀羊的能力，羊很高大根本不理会狼的挑衅，狼很快就发现了羊和羊脖子上的羊绳的关系，就咬着羊绳想把羊拉出来，无奈力小拉不动，人们就拿个牛头把羊吓了出来，狼很快就学会了这一点。只看了一次就知道了牛头和羊的关系，真的是聪明绝伦，难怪狼群可以不被驯服却照样活了下来。在技术层面，狼的聪明就在于它能很快发现事物之间的关联，并且充分利用这样的关联改变自己的行为，这是多么智慧的生灵，当然是一个有知识的行为。



建立企业生存和发展的根基，就是要建立一种适应环境的能力和-learning能力，而员工只是表现这种企业所有员工共有的这种企业基因的一个载体。这样，在环境相同的时候，所有员工的表现基本一致，实现了企业行为的稳定性和规范性，同时又具备了在变化环境下能发现事物之间关联的知识发现的能力，由此建立企业稳定发展的基础。这里的关键是，如何实现

在企业拥有员工之前就先拥有能力，如何承载这个能力呢？只有一个办法，这就是必须建立一个软件平台，把已有的对象、以及发现对象之间关联的能力，都根植在这个软件平台上。首先是企业的内部知识的梳理，最后进化到整个 web 网络，整个社会由此进入到知识时代。