

## 非结构化数据分析：大数据时代新价值

“大数据”是近两年来各界最热议的话题之一，和“大数据”相关的实践几乎成为所有行业在技术和商业两个方面最具创新意义的行动。在越来越多企业重视“大数据”价值的同时，新型数据的挑战也随之而来。

据IBM商业价值研究院和牛津大学赛德商学院共同发布的《分析：大数据在现实世界中的应用》显示，全球仅四分之一的受访者表示自己具备了分析高度非结构化数据的能力，而对大部分组织而言，掌握先进的非结构化数据分析能力仍是从“大数据”中获得价值的重大挑战。究竟“非结构化数据”与“大数据”有何关联？哪些创新技术能够应对新型数据的挑战？

### 非结构化数据：新价值和新挑战

IDC研究表明，数字领域当前存在着 1.8 万亿 GB 的数据。在这之中，除了传统数据库的结构化数据，还有像文件、文献、表单、影像等无以计量的非结构化数据。

什么是“非结构化数据”？相较于记录了生产、业务、交易和客户信息等的结构化数据，非结构化的信息涵盖了更为广泛的内容，包括了如合约、发票、书信与采购记录等营运内容；如文书处理、电子表格、简报档案与电子邮件等部门内容；如HTML与XML等格式信息的Web内容；以及如声音、影片、图形等媒体内容。

目前，非结构化数据的内容占据了当前数据海洋的80%，并将在2020年之前以44倍的速度迅猛增长。同时，因为非结构化数据的信息量和信息的重要程度很难被界定，分析成为了难点。如果说结构化数据用详实的方式记录了企业的生产交易活动，那么非结构化数据则是掌握企业命脉的关键内容，所反映的信息蕴含着诸多企业效益提高的机会。因此，只有解决非结构化数据的分析困难，才能有效挖掘这些数据背后的价值，克服逐渐攀升的数据量和复杂性对企业生产发展的重大阻碍，驱动企业价值提升。

### 企业内容管理：突围非结构化数据

针对非结构化数据所呈现的新趋势，IBM近年来不断加深在企业内容管理

(ECM) 领域的实践，充分发挥信息全文检索、自然语言处理、和行业深入理解三大突出能力，为越来越多的企业找到了非结构化数据管理和分析之道。

- **BJC Healthcare推进医疗研究：**和大部分医疗服务提供商一样，BJC Healthcare面临大数据时代的挑战。现有的医疗信息资源与实际操作严重脱节，分散在非结构化的临床记录和诊断报告中的历史数据几乎闲置，只有少部分研究人员偶尔查阅，其隐藏的研究信息因此无法得以充分利用。通过使用IBM 企业内容管理解决方案，BJC和大学研究人员得以分析非结构化的历史存档来回答以前无法解决的关键问题。比如，患者吸烟吗？如果吸烟，那么他吸烟的频率和烟龄是多少？如果已经戒烟，那么戒烟多长时间了？患者自己用过哪些药物？患者自己对病情是如何判断的？医院曾经对患者执行了哪些治疗过程？——通过这些问题的梳理，BJC现在能够更深入地了解医疗信息，揭示病历内容中存在的趋势和模式，向患者提供更有有效的医疗服务。
- **赫兹提升客户满意度：**赫兹国际汽车租赁公司是全球最大并广受欢迎的租车品牌，准确识别客户关于产品、服务和员工的满意度是赫兹的关键业务。通过使用IBM 企业内容管理解决方案，赫兹能够通过分析客户的调查记录更好地识别汽车和设备租赁业绩水平，确定客户的满意水平并进行必要的调整，从而进一步改善流程。此外，内容分析结果页能够支持赫兹推出针对性的营销活动，不断调整其产品和服务来提升客户的满意度。
- **北卡罗来纳州立大学识别商业机会：**北卡罗来纳州立大学认识到，大学的本质不仅仅是教育培训，而是以多种形式推进和传播知识。通过使用IBM企业内容管理解决方案，北卡罗来纳州立大学能够有效挖掘和分析庞大的数据，分析来自非结构化信息源的数千项内容，执行数据和文本分析，从而更好的识别哪些科研成果能够投入实际生产。通过分析所获取的洞察，它的技术转让办公室(Office of Technology Transfer)能够以更经济有效的方式将本大学研制的技术授权给企业。

以上仅是IBM企业内容管理（ECM）应对非结构化数据分析的几个案例，如今，ECM所带来的深刻洞察已经逐步成为金融、电信、保险、医疗、制造等行业的核心能力。此外，在多个增长市场ECM也扮演着重要角色。在智慧城市

（**Smarter City**）中，非结构化数据分析成为政府、公安、消防、民政等多个部门的协同运营奠定基础；在智慧商务（**Smarter Commerce**）中，非结构化数据分析为产品信息管理、销售合同管理、服务管理、市场推广行动管理提供支持；在社交商务（**Social Business**）中，非结构化数据分析成为网络协同，信息挖掘的前提。