Information Analysis Series

·软·科·学·研·究·方·法·系·列·

数据仓库和数据挖掘

苏新宁　杨建林　江念南　栗　湘　编著

信息分析丛书 | 丛书主编　包昌火　谢新洲

清华大学出版社

数据仓库和数据挖掘

信息分析丛书

丛书主编

包昌火 谢新洲

数据仓库和数据挖掘

苏新宁 杨建林 江念南 栗 湘 编著

清华大学出版社

北京

内 容 简 介

20 世纪 90 年代兴起的数据仓库和数据挖掘代表着信息序化和信息分析技术的重大进展。两者的结合，已成为人类处理和分析海量信息的有力武器。

本书在论述数据仓库和数据挖掘技术基本概念的基础上，系统和深入地剖析了数据仓库的模型，以数据仓库为应用平台的联机分析处理（OLAP）技术，以证券行业为对象的数据仓库的开发实例，数据库挖掘、文本挖掘、Web 挖掘、数据挖掘软件，以及数据挖掘的应用，尤其在竞争情报系统和客户关系管理中的应用，从而为了解和掌握数据仓库和数据挖掘技术提供了一个知识门户。

本书可供我国企业界、情报界、咨询界、教育界的信息分析、竞争情报、信息管理、知识管理、战略管理和软科学研究从业者的专业进修，以及高等院校师生教学和参考之用。

图书在版编目(CIP)数据

数据仓库和数据挖掘 / 苏新宁等编著. —北京：清华大学出版社，2006.4
（信息分析丛书）
ISBN 7-302-12648-8

Ⅰ. 数…　Ⅱ. 苏…　Ⅲ. ①数据库系统 ②数据采集　Ⅳ. ①TP311.13 ②TP274

中国版本图书馆 CIP 数据核字（2006）第 017812 号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**客户服务**：010-62776969
责任编辑：汪汉友
印 装 者：北京鑫海金澳胶印有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×260　**印张**：20　**字数**：483 千字
版　　次：2006 年 4 月第 1 版　2006 年 4 月第1 次印刷
书　　号：ISBN 7-302-12648-8/TP・8085
印　　数：1～3000
定　　价：29.00 元

序 言

PREFACE

——如何培养面向未来知识密集型经济从业者的信息分析能力[①]

在当今复杂多变的商业世界中，获得关于顾客、市场或是竞争者的信息，已经成了人们日常工作中司空见惯的活动，但其过程往往缺乏组织性和系统性。只有那些目标明确、思路清晰的组织和企业，才有可能更好地把握竞争态势，在今天的全球市场中握有胜算。我将这一搜集和使用市场信息的学科定义为竞争情报，而信息分析正是这一学科的核心。这套有关信息分析的综合性丛书，将为学习和掌握这一关键学科奠定必要的基础。

正如包昌火研究员所定义，信息分析（或称情报研究）是一类通过系统化的过程将信息转换为知识、情报和谋略的科学活动的统称。本系列丛书涉及实现这一转变的原始资料和方法。我认为最有意思的部分当属信息分析的应用过程。受过专业培训的业务人员应用信息分析方法后，能够为他们所在的组织带来竞争情报，从而进一步为该组织在特定的市场中带来竞争优势。在未来的全球经济中，搜集、分析和生产高效竞争情报的知识与技巧将成为关键的竞争力。

在今天的竞争情报领域中存在着诸多强大的变动力量或趋势，对于创造这样一种竞争情报能力非常有利。我认为下述三点尤为重要。

（1）世界范围内的经理人与老总们越来越意识到竞争情报的价值和用途。

（2）整体而言，商业全球化的图景越来越清晰，这就要求商家有效借助竞争情报开展竞争继而取得成功。

（3）具有先进的信息技术以及训练有素的信息技术职业人的增加会极大地增强竞争情报在未来所发挥的效力。

归根结底，对于信息分析和高级竞争情报能力的职业化应用，会激发商业企业学习并使用它们的兴趣，并给他们提供必要的工具，从而在明天的知识经济竞争中大获全胜。

---

① 丛书序言《如何培养面向未来知识密集型经济从业者的信息分析能力》的作者为简 P. 赫林（Jan P. Herring，美国前科学技术情报官、摩托罗拉公司前情报主管、赫林联合公司总裁、竞争情报学院创始人之一）。本序言由北京大学新闻与传播学院王宇博士生翻译，北京城市学院竞争情报研究所李艳博士审校。

为知识经济而时刻准备着！

彼得·德鲁克（Peter Drucker）曾经把新的商业世界描绘为一个“知识社会”，在这个社会中，知识是关键资源，而知识工作者则是劳动力中的主导群体。在当前的环境下，这些所谓知识工作者包括那些拥有相当职业知识的人群，比如说医生、工程师、律师、会计师，以及科学家。然而，德鲁克教授认为，在未来的十年里，知识工作者中最重要的新生力量将会出现在一个新群体中，称之为“知识技工（Knowledge Technologists）”。这些新的知识工作者包括计算机科学家、软件设计师、实验室技术员、助理律师以及各种各样的信息提供者，比如竞争情报从业者。

这些知识技工把与工作相关的知识和计算机技能结合起来。他们在工作中手脑并重，不过即使是所进行的体力劳动也需要以相当的理论知识为基础，而这些知识是通过适当的培训项目得来的。这套信息分析丛书将帮助那些在知识技术职业中工作的人们夯实基础。

这种新型的知识工作者对于未来的经济发展至关重要，然而，面向他们的正式培训项目和相关课程目前还不存在。当我问及德鲁克教授未来一代的知识工作者将在哪里接受培训的时候，他给我提示了一个范本：本地的两年制社区大学——许多精通技术的年轻人将在这里学习基础计算机知识及其应用。但问题在于，大多数人对于将如何以及在哪里应用这些新的知识却一无所知。

其中一些通晓计算机技能的青年学生将会继续进入四年全日制大学深造，学习计算机、商业管理，以及其他应用学科的知识。但是，上述这些商科学习和职业教育却很少在强化学生的计算机技能方面有所贡献。而且，目前可能还没有任何一个专业能够提供相应的教育培训以发展学生的信息分析能力。

这套有关信息分析的丛书内容丰富，论证深入，可以为商业管理、工程以及其他职业教育领域，例如医学研究等专业提供补充。

据我所知，**这套丛书是历史上第一部面向 21 世纪知识经济下知识工作者学习所需的教科书**。本丛书的逻辑构架既适合于普通的知识工作者，同样也适用于更加专业化的知识技术人员，特别是那些在竞争情报领域工作的人们。

接下来，对各卷的情况做一概览。

信息分析导论

本卷旨在为学生提供关于信息分析的入门知识，同时对其应用范围加以介绍。我尤为偏爱探讨信息分析未来趋势的那一章，因为它向读者展示了该学科不断发展的走势。

竞争情报导论

本卷全方位地探讨了信息分析在商业和竞争情报领域的应用。各章内容覆盖了竞争情报的全部组成部分，包括如何组织和管理一个竞争情报部门，及其在中国和国外的发展现状。本卷同样以对该领域未来趋势的探讨结束。

信息源和信息采集

有了前两卷一般性的介绍内容作为基础，从本卷开始，丛书逐步进入了对信息分析

基本领域进行的深入研究。首先就是相关信息的搜集或获取问题。这在当今这个信息密集的世界当中可绝非易事。来自全球范围的各种二手信息如洪水般涌来，令人难以招架。信息的复杂程度之高、语言载体之多以及数量之大早已远远超出了个人的驾驭能力。计算机和高级信息技术当然是必要的，但对于真正掌控和理解所有的可用信息还远远不够。定位准确得当的信息搜集、处理和人工编译工作都是不可或缺的。本卷的最后一章内容独到而恰当，它探讨的在以知识为基础的竞争环境中进行信息搜集的相关法律和道德问题将会为个人以及组织的行为带来极大的影响，因而，这一部分一定会博得政府和商业领导们的青睐。

信息分析的方法与技能

本卷探讨的是信息分析问题中的另一方面。在信息搜集和分析之间寻求恰当的平衡永远都是一个颇具挑战性的课题。竞争情报人员常常花了太多时间用于信息搜集，而用于信息分析的时间却少之又少。合理平衡的搜集工作通常是有一个结构优良的分析框架作为指导的，而且在搜集开始之间就已经建立了。这一卷的内容全面丰富，既涉及商业领域又涉及技术层面，而且还包含了在今天这个复杂的世界中应用广泛的模拟分析和情景分析。需要指出的是，在对分析方法的未来发展趋势进行分析的基础之上，这一卷还突出介绍了有关计算机分析技术的内容，这是最令人感兴趣的发展趋势之一。

专利信息采集和分析

在今天这个以高科技为核心的商业世界中，版权和专利是知识经济的一种重要的货币形式。就这一话题进行的分析对于我们理解版权和专利的参与者及其所代表的市场至关重要。其信息分析结论在技术和商业领域都有用武之地。此外，尽管没有特别指出，但在使用专利分析时所涉及的法律和道德问题也是一类重要的职业责任。

数据仓库和数据挖掘

尽管在一般人看来，这个问题并不是信息分析或竞争情报的主要议题，但辟专章对其加以讨论也是十分有用的。一直以来，数据挖掘关注的是顾客、市场和公司内部数据——而并没有将竞争对手数据纳入视野。而且这一过程主要是以数据为导向而非以文本为导向。这一卷对这两个层面都有所关注。值得一提的是，本卷还设专章对数据仓库和数据挖掘的应用进行了梳理。

信息分析和竞争情报软件

本卷的内容是非常到位和及时的。正如我在 2002 年欧洲竞争情报从业者协会年会所做题为“竞争情报的未来”的大会发言（随后发表在《竞争情报杂志》2003 年 3—4 月第 2 号第 6 卷）中所描述的那样，高级信息技术有助于企业成功并提高信息技术的有效性。尽管当时对竞争情报能力发展所作的预测已经遥指到 2012 年，但据我观察，创造这一未来的所有必要信息基础——包括软件和硬件——在今天就已经齐备了！然而遗憾的是，信息技术组织和公司利用高级信息技术进行信息分析的步伐颇显迟滞，这主要是因为其管理层还不太情愿在必要的安装和运行环节做出投入。此外，大多数竞争情报工

作人员也不愿意投入时间和精力来学习和应用这些新的软件操作方法。

另外一个问题就是一直以来缺乏关于这一主题的好的教科书。本卷将会有效地填补这一空白。到目前为止，本卷是市场上相关主题最完整的文本。辅之以恰当的课堂指导，本书将会极大地增进人们对竞争情报软件及其应用的了解和接受。

信息分析和竞争情报案例

案例研究法几乎一直以来都是所有商业管理相关专业的基本讲授方式。它把“现实世界”带入了课堂，这样学生们就可以把刚刚学到的方法和技巧用于实践。这种教育方法是现实经验所证明和认可了的。

不过，遗憾的是，能够用于教育目的的优秀竞争情报案例奇货可居（参见吉拉德和赫林，Gilad and Herring，《商业情报分析的艺术和科学》，JAI 出版社，1996 年）。作为信息分析系列丛书的一部分，这一卷特别用于支持和强化其他卷本中所探讨过的内容。本卷内容设计精妙，是对整个丛书的有益补充。

信息经济学与信息分析

第九卷所探讨的话题——经济学，特别适用于整个丛书的主题。信息以及信息分析在很大程度上依赖于其采集和应用的成本收益分析。更重要的是，信息分析的应用环节对于其价值的实现至关重要。比方说，在把信息分析应用于竞争情报的过程当中，人们总是要寻求某种商业优势和/或收益。我的一个老朋友罗伯特·斯蒂勒（Robert Steele）曾经一针见血地指出，“信息花钱——而情报挣钱”。所以信息分析在商业领域的应用拥有强大的经济推动力。

我对这一卷所覆盖话题的广度深表赞赏，它从基本经济理论入手，拓展至市场营销领域，甚至还涉及社会和军事问题。这些案例分析会极大地帮助学生理解信息分析各领域的应用和效果。

由这九卷所构成的信息分析系列丛书从本质上来看是对“商业世界”中有关情报资源和方法——及其应用——进行开发和使用的全部现有知识的汇编。17 世纪早期，弗朗西斯·培根在他的鸿篇巨著《新亚特兰蒂斯》（New Atlantis）中曾经描绘了如何通过对当时已有的情报资源进行整合运用从而为虚构中的国家谋求福祉的方法。从那以后，还没有第二个人能够觊觎完成如此浩大的工程。然而，在培根的时代，把这种知识确立下来并且代代相传的任务还只局限于少数特定的工匠以及他们所属的行会。时至今日，学术教育专家对于书籍的恰当使用已经使得该过程更加有效而广泛。这套《信息分析丛书》与此目标恰好是吻合的。

思考与总结

在我们审视这套独一无二的丛书所涉及内容的范围和广度的时候，我们清楚地看到，丛书的成功从根本上有赖于每本书的作者所做出的贡献和努力。整套丛书的价值毫无疑问是大于任何一卷书的单一价值。根据我在竞争情报领域 30 多年来摸爬滚打的经验，我可以毫不夸张地指出，**这项工程具有里程碑式的意义**。这套丛书的出版对于该书的最初倡导者和帮助它一步步变成现实的人们而言，实在是一种最大的褒扬和赞颂。

总之，这项具有前瞻性和巨大影响力的信息分析和竞争情报知识汇编工作，会给许多未来一代商业和竞争情报从业者提供重要的学习经验和途径。但可能更为重要的是，它给德鲁克教授口中的那些知识工作者们提供了一个教学的范本。接受过专门培训的未来一代信息和情报分析师们会使在不断发展的全球经济中的中国企业受益匪浅。这套丛书的价值不容低估。

简 P. 赫林（Jan P. Herring）

2005 年 5 月

前言 FOREWORD

1990 年由科技文献出版社出版、曾获全国优秀科技图书二等奖的《情报研究方法论》问世至今已经 15 年。15 年来，随着经济市场化、社会信息化和决策科学化的发展，我国情报研究的环境发生了重大的变化。《信息分析丛书》的编著可以看作在新世纪我国情报界为提升情报研究工作和咨询业的核心能力、加强开发信息资源的技术基础，以满足社会对日益增长的信息资源开发和信息分析能力培养的迫切需求所作的重要努力，是《情报研究方法论》在新形势下的“续集”。

一、情报研究学科

1956 年 2 月 11 日，我国第一个国防科技情报机构和 1956 年 10 月 15 日中国科学院科学情报研究所的成立以及尔后其他部委级和省市级等情报机构的建立，标志着我国一个条块结合、纵横交错，为科技发展和领导决策服务，在现代化建设中起耳目、尖兵和参谋作用的专职情报研究体系的形成。

1978 年以武衡为理事长的中国科技情报学会成立。1983 年第二届中国科技情报学会组建了情报研究专业委员会，由铁道部科技情报研究所何璧任主任委员。从此，作为我国科技情报两大主要工作之一的情报研究有了自己的学术研究和交流的组织和平台。这在国际上恐怕也是绝无仅有的。此后，中国科技情报研究所孙学琛和中国兵器工业集团第 210 研究所包昌火分别于 1988 年和 1993 年任第二届和第三届情报研究委员会主任。1994 年 1 月 28 日经中国科协批准、在情报研究专业委员会的基础上，成立了以包昌火为主任的情报研究暨竞争情报专业委员会，促进了情报研究和竞争情报的融合。1995 年 4 月 28 日，经中国科协批准、民政部登记，成立了以包昌火为理事长的中国科技情报学会竞争情报分会，兼行情报研究专业委员会的职能，对外亦称中国竞争情报研究会（Society of Competitive Intelligence of China，SCIC），从而推进了竞争情报在中国的崛起和发展。1998 年召开的第五届中国科技情报学会理事代表大会成立了名为信息研究与咨询专业委员会，中国科技信息研究所张钟为主任委员；2003 年召开的第六届理事代表大会成立了信息咨询专业委员会，中国人民大学信息资源管理学院卢小宾任主任委员。尽管称谓变迁，但情报研究这项学科建设的研究和发展一直是我国情报界和中国科技情报学会关注的重点之一。

例如，1987—1990 年，由当时的国防科工委下达的“情报研究量化分析”课题，阐明了情报研究学科的基本内容和主要特征；构筑了情报研究方法论的基本框架和科学体系；论述了现代情报研究的基本程序和主要方法；提出了情报研究方法论的技术手段和数学基础，为我国情报研究学科的建设和量化分析在我国情报界和企业界中的应用做出了贡献。

又例如，1991—1994 年，完成了由当时的国家科委和国防科工委联合下达的 “情报研究的国内外比较研究”课题，调查了国际情报研究机构的现状和特点；总结了我国情报研究机构的发展和经验；论述了情报研究学科的理论和发展；研究了竞争情报在我国的应用和前景；提出了对我国情报研究工作的认识和对策，为在市场经济条件下我国情报研究工作向咨询业方向转变提供了基本思路和改革框架，为国家科委和国防科工委制定情报研究工作的方针政策和发展规划提供了重要的依据和论证。

这些都表明，我国情报研究工作者为推进情报研究学科建设和发展作出了持续的努力和贡献，与国际同行相比，也毫不逊色。

二、信息和情报

信息（Information）和情报(Intelligence)是中国情报学的两个核心概念，是学科存在和发展的两大基石。正因为如此，对其探索和争论一直存在于我国情报界，并实际上已经形成了以高等院校一些学者为代表的 Information 学派和以专职情报研究机构一些专家为代表的 Intelligence 学派。这表明，信息和情报两大概念以及情报学是否应建立在 Information 和 Intelligence 两大基石上这一基本理论问题的重要性和复杂性。

笔者认为，信息和情报是两个既有区别又有联系的概念，不论中文还是英文都是如此。**信息不是物质，不是能量，是一个横跨于三个世界之中的信息世界**。它既可以是物质运动状态及其变化的反映，又可以是人类精神活动的产物，即包含本体论信息和认识论信息两大部分，如信号、图像、数据、事实、性状、情况、态势，以及显性和隐性知识等。**而情报则是对信息的解读、判断和分析，是人脑思维的产物**，具有对抗性、战略性、智能性、增值性和可行性(Actionable)等特点。信息和情报的联系在于：信息是情报的素材和载体，情报是信息的激活和升华；信息是原料，情报是产品。人们获取信息的目的是为了生产用于决策活动的情报和谋略，即 **Information 的 Intelligence 化**。它精辟地表述了情报和信息两者的区别和联系，如经判断的信息，经分析形成的报告等。一项重要的情报既可能是一条原始信息，也可能是一份报告，关键在于它是否满足用户决策的需要。曾任美国参议院情报特别委员会的民主党领袖、总统对外情报顾问委员会顾问舒尔斯基（Abram N. Shulsky）在《无声的战争——理解情报世界》中认为：“情报分析是指把搜集来的信息碎片转化为决策和军事指挥者可以使用的形式”(《美国军事情报理论著作评价》，张晓军，时事出版社，2005）的论述也佐证了笔者对信息转化为情报的理解。由此可见，信息的序化和转化，或者再聚焦一下，信息的获取和分析，即广义的 Information 的 Intelligence 化，是一切情报活动的基本任务，也应是我国情报学研究的核心问题。实际上，情报学并非起源于文献学和图书馆学，而应是起源于军事学和谋略学，起源于人类的情报活动和咨询活动，与人类的竞争和决策相伴相生，它们形影相随，而又若即若离。如盛行于我国春秋战国时期的兵法和韬略，因而孙子也被国际情报界誉

为“情报之父”。在当今信息过载、情报稀缺的时代，情报活动更显得迫切和需要。

几十年来，我国情报学研究重视文献资源，忽视人际资源；重视文献交流，忽视人际交流；重视信息技术路线，忽视社会经济路线；重视 Information，忽视 Intelligence，用 Information Science 来指导本属 Intelligence Study 的情报研究，试图解决植根于人际网络的情报活动，从而导致了我国情报学研究与情报活动相分离，以至于无视我国近五十年来情报研究工作的杰出贡献和在情报学学科体系中的重要地位，这种状况该到改弦易辙的时候了。

近五十年来，情报研究在我国科技进步、经济振兴和社会发展中起到了**国家竞争的杠杆、经济发展的先导和决策科学化的支撑**的重要作用(《情报研究的国内外比较研究总报告》，兵工情报研究报告，BQB94-0278，1994)，并对我国现代化建设做出了不可磨灭的贡献。大庆、葛洲坝和宝钢的建设，“两弹一星”、核潜艇、新型飞机和主战坦克等从无到有的发展，高新技术的研制以及众多新学科、新概念的引进，无不凝聚着我国情报人员的汗水和智慧。

匪夷所思的是，在我国情报学的学科体系中，居然找不到情报研究学科的位置，这在国家技术监督局 1992 年 11 月 1 号发布的《学科分类与代码》(GB/T 13745-92)得到了充分的反映。有情报心理学，竟无情报研究学。近 10 多年来在情报学理论研究中，越来越明显地存在着忽视情报工作(Intelligence Service)，专注信息工作(Information Service)，忽视决策咨询，专注信息管理，其结果势必使我国情报学远离我国情报工作的实践，远离政府和企业的决策活动，形成情报学与图书馆学的合流，丧失情报学的核心领地，漂流在信息世界的汪洋大海之中。

三、情报研究、信息分析和软科学研究

与信息和情报的含义一直存在着争论一样，我国情报界对情报研究(Intelligence Analysis & Synthesis 或 Intelligence Analysis)、信息分析(Information Analysis)和软科学研究(Soft Science Research)也一直是仁者见仁、智者见智。但这种认识上的差异大多不在内涵而系外延，是方向相同而非南辕北辙。

1990 年，在《情报研究方法论》一书对情报研究做过如下的表述：从根本上讲，**情报研究是根据特定需要，对情报信息进行定向选择和科学抽象的研究活动**，是情报工作和科技工作相结合的产物，是一类科学劳动的集合。所谓定向选择，就是根据特定需要进行的情报搜集和信息整序工作。所谓科学抽象，就是透过现象，揭示研究对象的本质、规律和联系的思维过程。定向选择和科学抽象的结果，必然会形成新的情报或情报集合。没有定向选择，就缺乏情报性，没有科学抽象，就缺乏研究性。情报性和研究性的结合，就形成了情报研究的基本特色。

此论述表明，情报研究的基本含义就是对信息的获取和分析，并对分析的基本方法作了高度的哲学概括，就是包含科学思维、文献计量、数理统计、专家调查和系统分析等在内的科学抽象，形成新的情报产品，因此是一种知识创造活动。书中并把课题选择、情报搜集、信息整序、成果表达和成果评价作为中国情报研究的基本程序。

1994 年，在《情报研究的国内外比较研究总报告》中，笔者等对情报研究又作了如下的描述：情报研究是根据社会用户的特定需求，以现代的信息技术和科学研究方法论

为主要手段，以社会信息的采集、选择、评价、分析和综合等系列化加工为基本过程，以形成新的、增值的情报产品，为不同层次的科学决策服务为主要目的的一类社会化的智能活动。

此描述表明，情报研究是一项以满足社会用户科学决策的特定需要而从事的一项专门化的智能活动，具有很强的目的性、政策性和智能性；强调运用现代信息技术和科学研究方法论对社会信息进行系列化的深度加工是情报研究的重要特征，个别的加工环节和缺乏深度的加工过程不应视为情报研究；最终成果应是形成新的增值的情报产品。它可以是一种情况、背景或判断，也可以是一种思想、建议或方案，能产生一定的社会、经济效益。因此，新颖性和可行性应是情报研究成果的灵魂。

1991 年 2 月 20 日国家科委发布的《国家科学技术情报发展政策》(中国科学技术蓝皮书第 6 号) 明确指出："情报研究又称情报分析，是科技情报工作的重要环节。加强情报研究为决策科学化提供可靠依据，推动科学技术的进步和经济发展，是这项工作的基本方针。"蓝皮书认为："**情报研究是对情报的深度加工，属思想库范畴**。"这为情报研究的性质、任务作出了政策性的规定，迄今对于理解情报研究的含义仍然具有重要的指导意义。

随着 1992 年 9 月全国科技情报工作会议将"情报"改称"信息"的行政干预、情报研究领域的扩大和社会信息化的发展，信息分析一词的应用逐渐普及开来。

笔者认为，所谓**信息分析是通过系统化过程将信息转化为知识、情报和谋略的一类科学活动的统称**，从数据挖掘、市场调查、竞争情报到软科学研究，形成了一条很宽的研究谱带。鉴于 Intelligence 兼具信息、情报、智能、谋略和能力的含义，因此从内涵上看，**信息分析的本质就是 Information 的 Intelligence 化**。这也就是笔者将信息分析和情报研究看作同义词的基本依据。两者的混用在国外亦不鲜见，如美国国防部 20 多个信息分析中心，多用 Information Analysis；美国中央情报局，多用 Intelligence Analysis。视角不同，称谓各异。视信息为始点和原料，对信息进行分析，称信息分析；视情报为终点和结果，把分析过程视作情报的生产过程，称情报研究或情报分析("竞争情报的崛起和发展，"《情报学进展第五卷》，国防工业出版社，2003)。但从应用的领域来看，当前信息分析已越出了情报界的范围，广泛地进入了咨询、商务、金融和电信等领域，扩展至国民经济的各行各业，从而实际上成为我国情报研究的主要称谓。

软科学兴起于 20 世纪 80 年代，在英语国家中 Soft Science 是指与"硬科学"相比"不那么科学"的一些研究领域，或认为是运用数学较少的那些研究领域。但在日本，软科学被政府科学技术厅定义为"建立在情报科学、行为科学和系统工程基础上的一种新的用于预测、计划、控制和评价的科学技术方法论"(《情报研究的国内外比较研究总报告》，兵工情报研究报告，BQB94-0278，1994)。

在我国，软科学被认为是一类研究社会组织和管理的学科的总称，主要包括系统科学、管理科学、未来学和技术经济学等，它以现代科学研究方法为手段，以阐明现代社会复杂的政策课题为目的，对包括科学技术和社会现象在内的广泛范围的对象进行跨学科的研究，为有关决策提供科学依据。

从发展上看，我国的情报研究侧重于情况研究，软科学研究侧重于对策研究，而把情况研究和对策研究结合起来正是 20 世纪 80 年代以来我国情报研究工作的重大发展，

而且许多部委级和省市级情报研究机构的任务和性质也逐渐与软科学研究机构类同，在统计上属软科学机构之列。因此，可以认为，情报研究是软科学研究的重要方面军，但又从事诸如跟踪研究、动态分析、学科评述、技术攻关和课题查新等一般不属于软科学研究范围之列的工作，因而情报研究又是比软科学研究更为广泛的一类研究活动。

由此可见，情报研究、信息分析和软科学研究都是以信息的采集、分析为基础，以现代信息技术和研究方法为手段，都是为用户的科学决策提供导向和支持，同属咨询业范畴；而情报研究 “软化”和软科学研究 “硬化”的发展，又使得两类内涵类同、重点有别、外延模糊的科学活动逐渐出现了融合的趋势，这也是本丛书定名为：软科学研究方法系列·《信息分析丛书》的基本缘由。

四、信息分析丛书

“开发信息资源，服务四化建设”是邓小平提出的我国现代化和信息化建设的指导方针，而信息资源开发的核心、信息通向应用的桥梁是信息分析，以此形成新的知识来源和决策依据，服务于各类管理活动和科学决策。因此，培养和造就一大批信息分析从业人员就成为我国信息化建设的历史任务和战略举措，是一项重大的知识创新工程。

为此，由我国一些著名的情报研究和竞争情报专家主编和编著、清华大学出版社出版和发行的《信息分析丛书》将为提升我国信息分析、软科学研究和咨询业的核心能力，加强信息资源开发的技术基础，顺应政府部门、情报机构、高等院校和企事业单位对信息分析教育、培训和进修的迫切需要做出贡献。

我国著名的软科学研究专家、中国软科学研究会理事长、全国人大副委员长成思危教授，美国前国家科学技术情报官、美国竞争情报奠基人之一、著名的情报分析大师简·赫林（Jan P. Herring）先生任丛书顾问；中国科技情报学会理事长、中国科技咨询业协会理事长、国务院石定环参事为编委会主任。

丛书由**《信息分析导论》、《竞争情报导论》、《信息源和信息采集》、《专利信息的采集和分析》、《信息分析方法与技能》、《数据仓库和数据挖掘》、《信息分析和竞争情报软件》、《信息分析和竞争情报案例》和《信息经济学与信息分析》**构成，共计三百多万字。全书由中国兵器工业集团第210研究所、北京大学、南京大学、北京城市学院、中国农科院经济研究所、南京理工大学、国家知识产权局、中国普天信息产业集团公司战略发展部、中国科学院文献情报中心、中国社会科学院文献中心、北京市科委软科学处、中国信息协会市场研究业分会等单位的学者和专家联袂完成，**北京市科学技术委员会、北京大学中国竞争情报和竞争力研究中心与中国兵器工业集团第210研究所为丛书的编著单位。**

丛书从国际视野和方法论的高度出发，系统地论述了作为软科学研究方法论基础和咨询业核心能力的、广泛地存在于人类智力活动中的信息分析的基本理论、发展现状、主要方法、关键技术、分析案例和未来发展，是迄今为止国际上首部关于信息分析的学术专著和系列丛书。正如简·赫林先生所言：“据我所知，这套丛书是历史上第一部面向21世纪知识经济条件下知识工作者培训所需要的教科书”。“根据我在竞争情报领域30多年来摸爬滚打的经验，我可毫不夸张的指出，这项工程具有里程碑式的意义”。

本丛书的基本特色是学术性、新颖性、实用性和可读性相结合，它反映了信息分析

领域自20世纪90年代以来的最新发展，包含了诸如战略与竞争分析、综合集成研讨厅、可视化分析、数据挖掘、战争游戏、数据仓库和人际情报网络等前沿内容，尤其是专利分析、分析软件和分析案例，在国际上也不多见，适合于我国情报界、教育界、企业界和咨询界的信息分析，以及软科学从业者教学、培训、进修和研究之用。

笔者对为本丛书的编著和出版做出贡献的学者、单位、顾问、编委和清华大学出版社表示诚挚的感谢。

包昌火　谢新洲

2005年10月20日

主 编 者 按

20 世纪 90 年代兴起的数据仓库和数据挖掘技术为人类应对信息爆炸、处理海量信息提供了科学和有效的手段，代表着信息序化和信息分析技术的重大发展。

数据仓库是一种多维化的信息组织技术，是面向主题的集成的、相对稳定的、随时更新的数据集合，用于分析型信息处理，为数据挖掘提供平台；数据挖掘是一种新兴的信息分析技术，能从大量的结构化和非结构化数据中提取隐含其中的有用信息和知识的过程，是当代知识发现的重要工具。数据仓库和数据挖掘的结合，则已成为人类处理和分析海量信息的得力武器。

为此，本书围绕着上述两大主题，从情报学和应用实践的视角，避免复杂的算法讲解，采用深入浅出的语言和案例，论述了数据仓库和数据挖掘这类新兴技术的基本理论、主要内容、关键技术和实际应用，以便为广大读者和从业者提供对这类计算机信息处理和分析技术的总体把握和应用知识。

全书共分 11 章，在论述数据仓库和数据挖掘技术基本概念的基础上，系统和深入地剖析了数据仓库的模型，以数据仓库为应用平台的联机分析处理（OLAP）技术，以证券行业为对象的数据仓库的开发实例，数据库挖掘、文本挖掘、Web 挖掘、数据挖掘软件，以及数据挖掘的应用，尤其在竞争情报系统和客户关系管理中的应用，从而为了解和掌握数据仓库和数据挖掘技术提供了一个知识门户。

本书是集体合作的成果，南京大学信息管理系博士生导师、南京大学中国社会科学研究评价中心副主任、南京大学信息技术开发研究所所长、《情报学报》编委、我国著名的信息处理专家苏新宁教授进行框架设计和全书审定，南京大学信息管理系杨建林博士参与框架设计、审定并撰写第 1、7、8、10、11 章，南京证券公司江念南博士撰写第 2、3、4、5 章和南京大学信息管理系粟湘博士撰写第 6 章和第 9 章。

作　者

2006 年 3 月

目录 CONTENTS