

21世纪高等学校规划教材 | 信息管理与信息系统



商务智能与 数据挖掘

蔡晓妍 张阳 李书琴 编著

清华大学出版社

21世纪高等学校规划教材 | 信息管理与信息系统

A decorative graphic consisting of a stylized vine with leaves and a circular swirl, rendered in a light gray color, positioned behind the main title.

商务智能与数据挖掘

蔡晓妍 张阳 李书琴 编著

清华大学出版社
北京

内 容 简 介

商务智能与数据挖掘是近年来企业信息化的热点内容,有着广阔的应用前景。本书共分 10 章,包括商务智能概述、商务智能中的核心技术、数据挖掘概述、分类分析、关联分析、聚类分析、异常检测、Web 挖掘技术、RFID 数据挖掘、数据挖掘在电子商务中的应用等内容;汇集了统计学、机器学习、数据库、人工智能等学科,具有多学科交叉、技术与管理融合等特点。

本书内容全面、案例丰富,适合作为计算机应用、软件工程、信息管理、电子商务和管理科学等相关专业的本科生和研究生教材,也可作为一些企事业单位、政府部门、研究机构等从事商务智能理论研究工作的相关人员的参考资料。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

商务智能与数据挖掘/蔡晓妍,张阳,李书琴编著. 北京:清华大学出版社,2016

21 世纪高等学校规划教材·信息管理与信息系统

ISBN 978-7-302-41741-5

I. ①商… II. ①蔡… ②张… ③李… III. ①电子商务—高等学校—教材 ②数据采集—高等学校—教材 IV. ①F713.36 ②TP274

中国版本图书馆 CIP 数据核字(2015)第 239598 号

责任编辑:郑寅堃 薛 阳

封面设计:傅瑞学

责任校对:梁 毅

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:14.75 字 数:367 千字

版 次:2016 年 4 月第 1 版 印 次:2016 年 4 月第 1 次印刷

印 数:1~1500

定 价:29.50 元

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上。精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展,顺应并符合21世纪教学发展的规律,代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版

社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。推出的特色精品教材包括:

(1) 21世纪高等学校规划教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 21世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 21世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。

(4) 21世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。

(5) 21世纪高等学校规划教材·信息管理与信息系统。

(6) 21世纪高等学校规划教材·财经管理与应用。

(7) 21世纪高等学校规划教材·电子商务。

(8) 21世纪高等学校规划教材·物联网。

清华大学出版社经过三十多年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail:weijj@tup.tsinghua.edu.cn

前言

在 2015 年新一代信息技术产业发展高峰论坛上,工业和信息化部副部长杨学山指出,信息技术产业已经在快速发展的轨道上走了几十年的时间,今后还将继续快速发展。目前,信息技术正在原来的轨道及新的发展方向上产生着重要变化,走向智能化阶段。企业要想在激烈的市场竞争中脱颖而出,成为市场的大赢家,就必须运用各种现代化的计算机技术、网络技术,打造出一套属于自己的决策系统,通过共享、整合和分析顾客、供应商等相关的数据,从而以最快的速度对瞬息万变的市场做出反应,以最有效的方式做出决策。

商务智能是终端用户查询和报告、联机分析处理、数据挖掘、数据仓库等软件工具的集合,使用某种数学算法对数据进行分析并发现规律,从而建立一种商务模型,提供给管理决策层进行模拟分析的全面解决方案。

数据挖掘技术在商务智能中的应用,已成为各行业、各部门信息化的必然趋势。现代企业的发展面临着巨大的挑战,国际市场环境日渐开放,同业竞争日趋激烈,客户要求越来越个性化,各类信息量正在迅速膨胀。数据仓库技术有效地把操作数据集成到统一的环境中,提供用户用于决策支持所需的当前和历史数据,数据挖掘技术利用这种分析工具在这些海量数据中发现模型和数据间的关系,能够帮助企业更深入、更容易地分析数据,从海量的数据中寻找对自己真正有用的信息资源。商务智能通过对大量数据进行自动加工、处理和分折,实现数据向信息、信息向知识的转换,并将知识应用于决策。

本书讨论了商务智能基础,主要介绍商务智能的概念、产生与发展、价值以及目前的主流商务智能产品,使读者对商务智能有一个概要的认识;论述了数据仓库和在线分析处理、商务智能体系架构和数据挖掘等核心技术;详细介绍了数据挖掘的具体方法,即分类分析、关联分析、聚类分析和异常检测。在此基础上,讨论了商务智能与数据挖掘在 Web 数据挖掘、RFID 数据挖掘、电子商务等领域的应用。

本书由蔡晓妍任主编,各章编写分工如下:蔡晓妍编写第 1、10 章;李书琴编写第 7 章;张阳编写第 3 章;杨黎斌编写第 9 章;李梅编写第 5 章;张晓婷编写第 8 章;梁春泉编写第 6 章;朱珊珊编写第 4 章;杨丽丽编写第 2 章。蔡晓妍和杨黎斌负责全书的策划、大纲的制定和统纂工作。

本书在编写过程中,参考了国内外许多公开发表的相关资料,在此对所涉及的各位专家、学者表示诚挚的感谢。研究生许晶晶、由文浩、李飞、申昌、张嘉慧、韩亚敏和许允波同学分别对本书进行了校对并提出了宝贵的建议,特表示感谢。由于编写时间紧迫,加之编者理论水平 and 实践经验有限,书中难免有不当和疏漏之处,恳请广大读者批评和指正。

编者

2015 年 8 月

目 录

第 1 章 商务智能概述	1
1.1 商务智能的概念	1
1.1.1 数据、信息与知识	1
1.1.2 商务智能的定义	3
1.1.3 商务智能的特点	4
1.1.4 商务智能的过程	5
1.2 商业决策需要商务智能	5
1.2.1 管理就是决策	5
1.2.2 决策需要信息和知识	6
1.2.3 智能型企业	6
1.2.4 商务智能支持商业决策	6
1.2.5 新一代的决策支持系统	7
1.3 商务智能的产生与发展	8
1.3.1 商务智能的产生和发展过程	9
1.3.2 商务智能与其他系统的关系	10
1.4 商务智能的价值	12
1.5 商务智能系统的功能	14
1.6 主流商务智能产品	16
1.6.1 主流商务智能产品简介	16
1.6.2 商务智能的抉择	17
1.7 商务智能的未来发展趋势	18
1.8 商务智能的应用	19
1.9 本章小结	26
习题	26
第 2 章 商务智能中的核心技术	27
2.1 数据仓库	27
2.1.1 数据仓库的产生与发展	27
2.1.2 数据仓库的概念与特征	28
2.1.3 ETL	29
2.1.4 数据集市	31
2.1.5 数据仓库的数据组织	32

2.1.6	数据仓库的体系结构	32
2.1.7	数据仓库的开发步骤	33
2.2	在线分析处理	36
2.2.1	OLAP 简介	36
2.2.2	OLAP 的定义和相关概念	37
2.2.3	OLAP 与 OLTP 的区别	38
2.2.4	OLAP 的分类	39
2.2.5	OLAP 多维数据分析	40
2.3	商务智能体系结构	41
2.3.1	商务智能系统的组成	41
2.3.2	商务智能的架构体系	43
2.4	本章小结	44
	习题	44
第 3 章	数据挖掘概述	45
3.1	数据挖掘的起源与发展	45
3.1.1	数据挖掘的起源	45
3.1.2	数据挖掘的发展	46
3.2	数据挖掘所要解决的问题	47
3.3	数据挖掘的定义	48
3.4	数据挖掘的过程	49
3.5	数据挖掘系统	52
3.5.1	数据挖掘系统的分类	52
3.5.2	数据挖掘系统的发展	53
3.6	数据挖掘的功能和方法	54
3.6.1	数据挖掘的功能	54
3.6.2	数据挖掘的方法	55
3.7	数据挖掘的典型应用领域	58
3.8	数据挖掘的发展趋势	59
3.9	本章小结	61
	习题	61
第 4 章	分类分析	62
4.1	预备知识	62
4.2	解决分类问题的一般方法	64
4.3	分类算法	65
4.3.1	贝叶斯分类器	65
4.3.2	决策树	74
4.3.3	支持向量机	81

4.3.4	BP 神经网络	82
4.3.5	其他分类算法	86
4.4	评估分类器的性能	87
4.4.1	保持方法	88
4.4.2	随机二次抽样	88
4.4.3	交叉验证	88
4.4.4	自助法	89
4.5	本章小结	89
	习题	89
第 5 章	关联分析	92
5.1	引言	92
5.2	基本概念	93
5.3	关联规则的种类	95
5.4	关联规则的研究现状	96
5.5	关联规则挖掘算法	97
5.5.1	Apriori 算法	97
5.5.2	FP 增长算法	103
5.5.3	其他关联规则挖掘算法	106
5.6	改善关联规则挖掘质量问题	107
5.6.1	用户主观层面	107
5.6.2	系统客观层面	108
5.7	约束数据挖掘问题	108
5.8	本章小结	109
	习题	110
第 6 章	聚类分析	112
6.1	聚类的概念	112
6.1.1	聚类概念及应用	112
6.1.2	聚类算法要求	113
6.1.3	聚类技术类型划分	114
6.2	聚类分析的统计量	116
6.2.1	模型定义	116
6.2.2	相似性度量	116
6.3	常用聚类算法	121
6.3.1	k 均值算法	121
6.3.2	k -medoids 算法	123
6.3.3	凝聚层次聚类算法	124
6.3.4	DBSCAN 算法	129

6.3.5	STING 算法	131
6.3.6	CLIQUE 算法	132
6.4	簇评估	133
6.4.1	概述	134
6.4.2	非监督簇评估:使用凝聚度和分离度	135
6.4.3	非监督簇评估:使用邻近距离矩阵	138
6.4.4	层次聚类的非监督评估	139
6.4.5	确定正确的簇个数	139
6.4.6	聚类趋势	140
6.4.7	簇有效性的监督度量	141
6.5	与分类比较	143
6.6	本章小结	144
	习题	144
第 7 章	异常检测	146
7.1	预备知识	147
7.1.1	异常的原因	147
7.1.2	异常检测方法	147
7.1.3	类标号的使用	148
7.1.4	问题	148
7.2	统计方法	149
7.2.1	检测一元正态分布中的离群点	150
7.2.2	多元正态分布的离群点	150
7.2.3	异常检测的混合模型方法	151
7.2.4	优点与缺点	152
7.3	基于近邻度的离群点检测	152
7.4	基于密度的离群点检测	152
7.4.1	使用相对密度的离群点检测	153
7.4.2	优点与缺点	153
7.5	基于聚类的技术	154
7.5.1	评估对象属于簇的程度	154
7.5.2	离群点对初始聚类的影响	155
7.5.3	使用簇的个数	155
7.5.4	优点与缺点	155
7.6	本章小结	155
	习题	156
第 8 章	Web 挖掘技术	157
8.1	Web 数据挖掘概述	157

8.1.1	Web 数据挖掘的概念	157
8.1.2	Web 数据挖掘的特点	158
8.1.3	Web 数据挖掘的处理流程	158
8.1.4	Web 数据挖掘与信息检索、信息抽取的区别	159
8.2	Web 数据挖掘分类	160
8.2.1	Web 内容挖掘概述	160
8.2.2	Web 结构挖掘概述	161
8.2.3	Web 使用挖掘概述	161
8.3	Web 内容挖掘	162
8.3.1	特征提取和特征表示	162
8.3.2	自动摘要	163
8.3.3	文本分类	163
8.3.4	文本聚类	164
8.4	Web 结构挖掘	164
8.4.1	超链和页面内容的关系	165
8.4.2	不同挖掘阶段的分析	165
8.4.3	PageRank	166
8.4.4	HITS	172
8.4.5	两种算法的比较	176
8.4.6	Web 结构挖掘应用	176
8.5	Web 使用挖掘	177
8.5.1	Web 使用挖掘数据预处理	177
8.5.2	Web 使用挖掘模式发现	180
8.5.3	Web 使用挖掘模式分析	183
8.5.4	Web 使用挖掘模式应用	183
8.6	本章小结	185
	习题	185
第 9 章	RFID 数据挖掘	186
9.1	RFID 数据挖掘的发展	186
9.2	RFID 数据挖掘的作用	187
9.3	RFID 数据分析的典型应用	187
9.3.1	零售仓储	188
9.3.2	通关检查	189
9.3.3	运输管理	189
9.3.4	医疗管理	191
9.3.5	其他应用	191
9.4	本章小结	192
	习题	192

第 10 章 数据挖掘在电子商务中的应用	193
10.1 电子商务中数据挖掘的发展状况	193
10.2 电子商务中数据挖掘的特点	195
10.2.1 电子商务中数据挖掘的数据源	195
10.2.2 路径分析	197
10.2.3 电子商务中数据挖掘的体系结构	197
10.3 网站客户分群	198
10.4 优化网站结构	204
10.4.1 网站结构优化手段	205
10.4.2 网站结构优化模型	205
10.5 智能搜索引擎	207
10.5.1 传统搜索引擎的特点	207
10.5.2 智能搜索引擎的特点	209
10.5.3 网络爬虫	210
10.5.4 智能搜索引擎的技术与发展	214
10.6 客户关系管理	216
10.7 网络主动防御	219
10.8 本章小结	220
习题	221
参考文献	222

商务智能概述

随着世界经济全球化的迅猛发展,产业国际化的趋势不断加强,企业必须能够在瞬息万变的环境下及时做出反应。为了迎接市场的挑战,企业需要对市场有准确的把握,分析顾客的消费趋势,找出企业经营中出现的问题,加强与供应链合作伙伴的关系,挖掘新的商业机会,并能够对未来进行预测。随着企业信息化的发展,各种应用系统产生的数据量平均每18个月就翻一番。如何充分利用这些数据资产,挖掘出决策者需要的信息,做出高质量的决策是企业管理者需要考虑的问题。近年来,数据集成、数据分析、大容量数据存储与并行处理等技术不断成熟,成本不断下降,企业各种应用软件积累了大量数据,这些因素促进了商务智能的发展。商务智能(Business Intelligence, BI)可以将各种数据及时地转换为支持决策的信息和知识,帮助企业管理者了解顾客的需求、消费习惯,预测市场的变化趋势以及行业的整体发展方向,进行有效的决策,从而在竞争中占据有利地位。

1.1 商务智能的概念

商务智能越来越受到学术界和产业界的青睐,逐渐成为目前国内外企业界和软件开发界备受关注的研究热点。作为一项新兴的技术,在过去的十多年间,围绕商务智能的理论、方法、技术等的应用已经取得了许多令人瞩目的成就。

由于企业积累的这些海量数据不仅没能给企业带来财富,相反却使得企业淹没于数据之中,形成一个个信息孤岛和数据坟墓。企业面临着由于数据库变得越来越庞大而带来的对数据管理的困难。如何充分利用这些数据,为企业的经营决策服务?这就需要一种合适的数据处理和数据分析工具。

1.1.1 数据、信息与知识

在信息化时代,企业的智能已成为其生存之本。智能企业的发展会经历以下几个阶段:数据和应用集成、把数据转化为信息、信息转化成知识以及知识转化为行动。

1. 数据

数据是用来记录、描述和识别事物的按一定规则排列组合的物理符号,是一组表述数量、行动和目标的非随机的可鉴别的符号,是客观事物的属性、数量、位置及其相互关系等的抽象表示,以适合于用人工或自然的方式进行保存、传递和处理。它既可以是数字、文字、图

形、图像、声音或者味道,也可以是计算机代码。在计算机科学中,数据是指所有能输入到计算机中具有一定意义的数字、字母、符号和模拟量等并能够被计算机程序处理的符号的介质的总称,是计算机能够识别的二进制数的形式。

数据本身是孤立的、互不关联的客观事实、文字、数字和符号,没有上下文和解释。数据表达的仅仅是一个描述,如 20140112,我们只知道这是一个数字,或者可以看做是日期,对于这个数字来说,它就是数字,不表示别的任何含义。

数据是用属性描述的,属性也称变量、特征、字段或维。数据经过处理仍然是数据,只有经过解释,数据才有意义,才能成为信息。

2. 信息

信息是指人们对数据进行系统地收集、整理、管理和分析的结果,是经过一系列的提炼、加工和集成后的数据。信息是对客观世界各种事物特征的反映。数据是信息的符号表示,或称载体,数据不经加工只是一种原始材料,其价值只是在于记录了客观数据的事实。信息是数据的内涵,是对数据的解释。信息可以是完整的,也可以是片段的;可以是关于过去的,或者关于现在的,也可以是涉及未来的。“目前天气很热,气温高达 35℃”,这条信息描述的是现在的天气状况;“参考过去连续三年的气温记录,每年这一天的历史温度都高于 37℃”,这是关于过去的信息;“根据这两天的信息预测明天的气温至少为 37℃”,这是涉及未来的信息。尽管明天是高温天气是有可能的,甚至是必然的,但这种预测未来的信息多少会带有不确定性,为了减少不确定性,提高置信度,必须对信息进行提炼、加工和集成。

3. 知识

所谓知识,就它反映的内容而言,是客观事物的属性与联系的反映,是客观世界在人脑中的相对正确的反映。就它反映的活动形式而言,有时表现为主题对事物的感性直觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识。知识是在实践活动中获得的关于世界的最本质的认识,是对信息的提炼、比较、挖掘、分析、概括、判断和推论。

一般而言知识具有共享性、传递性、非损耗性(可以反复使用,其价值不会减小)及再生性等特点。

按知识的复杂性可将知识划分为显性(Explicit)知识和隐性(Tacit)知识,它是知识最基本和最重要的划分结构。显性知识是用系统、正式的语言传递的知识,可以编码和度量,可以清晰地表达出来,易于传播,可以在人与人之间进行直接的交流,通常以语言文字(如书籍、文件、网页、电子邮件等)形式存在。显性知识的处理可以用计算机实现。隐性知识是存在于人脑中的、非结构化的、与特定语境相关的知识,很难编码和度量。隐性知识是人们在实践中不断摸索和反复体验形成的,通常以直觉、价值观、推断、经验、技能等形式表现出来。它难以表述,但却是个人能力的直接表现且更为宝贵。隐性知识的处理只能通过人脑实现,一般要通过言传身教和师传徒等形式传播。

数据、信息和知识之间的关系为:从数据中提取信息,从信息中挖掘知识,如图 1-1 所示。



数据-信息-知识
数据：是信息和知识的符号表示
信息：数据中的内涵意义
知识：是一套具有前因后果关系的信息，是人们在长期的实践中总结出来的正确内容

图 1-1 数据、信息和知识的区别

1.1.2 商务智能的定义

1989年美国加特纳公司的分析师 Howard Dresner 首次创造出“商务智能”这个名词。商务智能可以定义为一组数学模型和分析方法，它们系统地开发可以利用的数据，以便检索对支持复杂决策过程有用的信息和知识。出于帮助企业所有者打造更好的商业的目的，商务智能是涉及收集、提供、存取及分析数据的一大类别的应用与技术。

商务智能从产生以来一直发展较快，但目前还不成熟，企业界和学术界对商务智能存在着或多或少不同的理解，这里分别列举几个比较全面的最新的定义。

IBM 对 BI 的定义是：“商务智能是一系列技术支持的简化信息收集、分析的策略集合。通过使用企业的数据资产来制定更好的商务决策。企业的决策人员以数据仓库为基础，通过各种查询分析工具、联机分析处理或者是数据挖掘加上决策人员的行业知识，从数据仓库中获得有利的信息，进而帮助企业提高利润，增加生产力和竞争力。”

Business Object 公司对 BI 的定义是：“商务智能是一种基于大量数据的信息提炼的过程，这个过程与知识共享和知识创造紧密结合，完成了从信息到知识的转变，最终为商家提供网络时代的竞争优势和实实在在的利润。”

Microsoft 公司对 BI 的定义是：“商务智能是任何尝试获取、分析企业数据以便更清楚地了解市场和顾客，改进企业流程，更有效地参与竞争的过程。”

IDC 公司对 BI 的定义是：“商务智能是下列软件工具的集合：终端用户查询和报告工具、在线分析处理工具、数据挖掘软件、数据集市、数据仓库产品和主管信息系统。”

Oracle 公司对 BI 的定义是：“商务智能是一种商务战略，能够持续不断地对企业经营理念、组织机构和业务流程进行重组，实现以顾客为中心的自动化管理。”

Data Warehouse Institute 对 BI 的定义是：“商务智能是把数据转换成知识并把知识应用到商业运营的一个过程。”

商务智能专家王茁在总结了商务智能的众多版本之后给商务智能下了如下定义：“商务智能是企业利用现代信息技术收集、管理和分析结构化和非结构化的商务数据和信息，创造和积累商务知识和见解，改善商务决策水平，采取有效的商务行动，完善各种商务流程，提升各方面商务绩效，增强综合竞争力的智慧和能力。”

利用现代信息技术——这是这一定义中的关键之一。现代信息技术的发展催生了信息

经济和信息社会,在这一新型的经济和社会形态中,信息的爆炸式增长又产生了对能够处理和
控制信息的技术的强烈需求,商务智能正是新的信息技术在商务分析中的有效应用。

总结上述观点,商务智能是融合了先进信息技术与创新管理理念的结合体,集成了企业
内外的数据,进行加工并从中提取能够创造商业价值的信息,面向企业战略并服务于管理
层、业务层,指导企业经营决策,提升企业竞争力,涉及企业战略、管理思想、业务整合和技术
体系等层面,促进信息到知识再到利润的转变,从而实现更好的绩效。事实上,商务智能应
用的核心不在其功能,而在于对业务的优化,IBM 公司更强调数据集成和数据分析基础
上的业务分析和优化(Business Analytics and Optimization,BAO)。目前,商务智能的应用已
延伸到了非商业领域,政府和教育部门等也成为商务智能的应用领域。

1.1.3 商务智能的特点

商务智能具有以下主要特点,了解这些特点有助于更好地理解商务智能的内涵。

1. 商务智能服务企业战略

商务智能能够对企业的内外部数据进行分析,支持企业战略管理。哈佛商学院的迈克尔·波特博士在 *Harvard Business Review* 的文章中把战略分为三个方面:定位、取舍和配称(各项运营活动之间如何关联),而商务智能可以通过数据分析帮助企业对这些方面进行规划。

2. 商务智能提升企业绩效

商务智能更多的是可以解决管理问题。通过商务智能能从企业多年运营的数据中,挖掘有效的模式辅助管理决策。随着商务智能应用的发展,商务智能离业务越来越近。商务智能在企业绩效管理扮演着重要的角色,而商务智能相关的产品在管理角色和方法、管理职能和过程等方面烙印渐深,并且融合了越来越多的企业管理的理念。例如,Business Objects(SAP)在 2007 年“商业智能 点亮明天”商务智能解决方案研讨会展示的绩效管理套件中包含了管理仪表盘、计分卡等工具。这些工具不再是色彩和图形的结合,而是包含了大量的业务逻辑关系和线性规划的运算模型。

3. 商务智能是“数据炼油厂”

商务智能可以看做是“数据炼油厂”,根据业务需要收集数据,并进行提炼和加工,最终产生对企业有价值的知识,提高企业的绩效。商务智能需要整合企业的业务系统数据,从而保证足够的“原料补给”。商务智能对 ERP、CRM 和 SCM 等业务系统中生成的运营数据进行分析,并给出报告,帮助管理者认识企业和市场的现状,预测发展趋势,做出正确的决策。

4. 商务智能是多项技术的综合应用

随着信息化的发展,商务智能已成为企业充分利用数据资产的重要方法。它从不同的数据源中提取有用的数据,通过数据仓库、在线分析处理和数据挖掘等技术实现企业的决策、考核、分析有机结合和量化以达到为企业提供经营管理、决策支持的目的。最新的商务智能还涉及其他一些新技术,例如内存中的分析处理、面向服务的软件架构(Service

Oriented Architecture, SOA)、文本挖掘和元数据存储等。商务智能在这些技术的支持下,发现数据背后隐藏的商机或威胁,获得洞察力,了解企业和市场的现状,把握趋势,识别异常情况,理解企业业务的推动力量,认清正在对企业的业务产生影响的行为及影响程度。

5. 商务智能用户的多样性

商务智能服务于各类企业决策者。传统应用中,商务智能主要支持中、高层管理人员决策。目前,商务智能平台的用户包括一线的业务人员、各级管理者,甚至外部的顾客和商业伙伴。这是因为业务经营决策的范围发生了扩展,包括操作层、战术层和战略层的决策。

1.1.4 商务智能的过程

商务智能的过程是:首先需要准备正确可用的数据,其次要将这些数据转化为有价值的信息,再用于指导商业实践(智慧)。该过程包括了数据抽取、分析和挖掘三个主要环节,分别由数据仓库(DW)、在线分析处理(OLAP)、数据挖掘(DM)技术来完成。

DW是商务智能的基础和核心,存储按照商务智能的要求重新组织来自业务系统的数据。

OLAP和DM在DW的基础上进行分析,提供给最终用户灵活自主的信息访问途径、丰富的数据分析和报表功能。

从系统的观点来看,商务智能中信息处理的过程可以归结为以下三个子过程:数据获取和管理、数据整理和分析、信息展现和表示。

(1) 数据获取和管理:从不同的数据源获取有用的数据,对数据进行清理以保证数据的正确性,将数据经过转换、重构后存入数据仓库或数据集市(这时数据变为信息)。

(2) 数据整理和分析:通过合适的查询和分析工具、数据挖掘工具,对信息进行处理(这时信息变为辅助决策的知识)。

(3) 信息展现和表示:将管理与决策所需要的知识呈现于用户面前,支持管理与决策。

1.2 商业决策需要商务智能

1.2.1 管理就是决策

管理学家西蒙指出:管理就是决策。决策是企业管理的核心,贯穿管理的全过程。根据决策活动特点,企业管理可分为战略层、中间管理层和运营层三个层次。企业的各个层次都需要决策。

(1) 战略层是企业决策中的最高层,负责管理、控制和协调整个企业的正常运行。厂址选择、资金分配计划、管理体制确定等都是战略层的决策范围。

(2) 中间管理层处于企业决策的中间层,主要包括销售、财务、生产、人力资源等部门,是实施战略层计划、管理控制运营层业务的关键所在,例如作业计划、作业调度、广告部署等。又如,某大型企业的区域销售主管把工作重心放在正确推断业务的发展方向上,用以确定该地区未来的业务重点。