

● 主 编 / 陈 荣 霍丽萍
● 副主编 / 朱世琴 曾 媛 严素梅

INFORMATION RETRIEVAL & CASE STUDIES

信息检索与 案例研究

● 主 编 / 陈 荣 霍丽萍
● 副主编 / 朱世琴 曾 媛 严素梅

INFORMATION RETRIEVAL & CASE STUDIES

信息检索与 案例研究



华东理工大学出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

· 上海 ·

图书在版编目(CIP)数据

信息检索与案例研究/陈荣,霍丽萍主编. —上海:
华东理工大学出版社,2015.4
ISBN 978-7-5628-4176-0

I. ①信… II. ①陈… ②霍… III. ①情报检索—研究
IV. ①G252.7

中国版本图书馆 CIP 数据核字(2015)第 037979 号

内 容 提 要

全书共分为 6 章,第 1 章为信息检索概念,第 2 章为信息收集方法,第 3 章为信息获取误差分析,第 4 章为信息分析与知识发现,第 5 章为信息管理与知识提炼,第 6 章为案例剖析。

本书可作为高等院校各专业本科生及研究生文献检索课程用教材,也可供广大科研人员和高校教师获取科研信息时参考。

信息检索与案例研究

主 编 / 陈 荣 霍丽萍

副 主 编 / 朱世琴 曾 媛 严素梅

责任编辑 / 焦婧茹

责任校对 / 李 晔

封面设计 / 裘幼华

出版发行 / 华东理工大学出版社有限公司

地 址:上海市梅陇路 130 号,200237

电 话:(021)64250306(营销部)

(021)64252712(编辑室)

传 真:(021)64252707

网 址:press.ecust.edu.cn

印 刷 / 常熟市华顺印刷有限公司

开 本 / 787mm×1092mm 1/16

印 张 / 16.25

字 数 / 402 千字

版 次 / 2015 年 4 月第 1 版

印 次 / 2015 年 4 月第 1 次

书 号 / ISBN 978-7-5628-4176-0

定 价 / 38.00 元

联系我们:电子邮箱 press@ecust.edu.cn

官方微博 e.weibo.com/ecustpress

淘宝官网 <http://shop61951206.taobao.com>

华东理工大学出版社



请扫描进入手机淘宝网店

前 言

早在 2007 年 12 月 17 日,全国工程硕士专业学位教育指导委员会发出《关于加强工程硕士“信息检索”课程建设的通知》(指导委[2007]13 号),将“信息检索”课程增列为工程硕士研究生的公共必修课。实际上,华东理工大学早在 1998 年就将“文献检索”课程列为工程硕士的必修课程。目前,文献检索课程不仅是华东理工大学工程硕士研究生的必修课程,而且在 2010 年研究生院将该课程列为信息素养平台课程。

与现有的文献检索教材相比,本书在框架和内容上有了较大的创新,以提高学生的信息素养为目的,提升学生在解决检索过程中碰到的实际问题的能力为目标,从科研过程的需求出发,以案例形式,对信息的检索及获取等进行了充分的介绍。本书的宗旨是加强实用性,希望每个老师和同学都能从书中找到自己需要的内容。

本书的第 1 章由朱世琴编写;第 2 章由陈荣和霍丽萍编写;第 3 章由曾媛编写;第 4 章由霍丽萍编写;第 5 章由严素梅编写;第 6 章和附录由陈荣、霍丽萍、曾媛、朱世琴、严素梅、刘云编写。唐永林、孙济庆和吉久明在本书的策划、编写等整个过程中均给予了耐心、细致、具体的指导,在此表示衷心的感谢!同时也非常感谢华东理工大学图书馆情报部的各位老师在本书的编写过程中所给予的关心和支持!

由于编者水平有限,书中难免有疏漏之处,敬请读者多提宝贵意见。

编者

目 录

第1章 信息检索概念	1
1.1 信息检索	1
1.1.1 信息检索的含义	1
1.1.2 信息检索的意义	1
1.2 信息的类型	12
1.2.1 从出版形式的角度划分	12
1.2.2 从加工层次的角度划分	17
1.2.3 从文献载体形式的角度划分	18
1.3 信息检索的发展趋势	19
1.3.1 信息检索的智能化	19
1.3.2 信息检索的可视化	20
1.3.3 信息检索的个性化	20
1.3.4 一站式信息检索	21
1.4 本章小结	22
第2章 数字信息收集方法	23
2.1 信息收集途径	23
2.1.1 搜索引擎	23
2.1.2 发现服务系统	27
2.1.3 信息数据库	29
2.1.4 开放获取	29
2.2 信息检索技术	36
2.2.1 布尔逻辑检索	36
2.2.2 字段检索	38
2.2.3 加权检索	38
2.2.4 引文检索	39
2.3 国内外主要数据库	40
2.3.1 主要的中文数据库	40
2.3.2 主要的外文数据库	41

2.3.3 专利数据库	45
2.3.4 数据库、检索技术及检索实例一览表	51
2.4 获取全文的途径	52
2.4.1 免费的学术搜索	52
2.4.2 图书馆主页链接	57
2.5 本章小结	62
第3章 信息获取误差分析	63
3.1 信息获取误差分析	63
3.1.1 检索结果过多或过少	63
3.1.2 检索结果与检索目的不符	68
3.1.3 中外文检索结果的差异	71
3.2 减少信息获取误差的方法	73
3.2.1 提高查全率的主要因素	74
3.2.2 提高查准率的主要因素	74
3.3 本章小结	74
第4章 信息分析与知识发现	75
4.1 基于文献特征的知识发现	76
4.1.1 基于文献特征的知识发现途径	76
4.1.2 基于文献特征的知识发现方法及工具	77
4.2 基于文献关系的知识发现	89
4.2.1 基于非相关文献的知识发现	89
4.2.2 基于相关文献的知识发现	94
4.3 本章小结	100
第5章 信息管理与知识提炼	101
5.1 文献管理软件——EndNote	101
5.1.1 EndNote 简介	101
5.1.2 EndNote 的基本功能	101
5.1.3 个人数据库的建立	102
5.1.4 个人数据库的应用	114
5.2 文献综述	118
5.2.1 文献综述的意义	118
5.2.2 文献综述的撰写	119
5.2.3 文献综述小结	122

5.3 期刊评价与论文投稿	123
5.3.1 期刊评价	123
5.3.2 论文投稿	125
5.3.3 科研选题	130
5.4 知识提炼——专利分析	131
5.4.1 专利分析方法概述	132
5.4.2 专利分析过程	132
5.4.3 专利分析工具的主要功能	133
5.4.4 国内外常用专利分析工具比较	134
5.4.5 专利分析的应用	136
5.4.6 专利分析案例	136
第6章 案例剖析	141
6.1 商学案例	141
6.1.1 网络商城营销策略创新研究	141
6.1.2 资本结构与公司业绩相关关系实证分析	145
6.2 文科案例	149
6.3 工学案例	154
6.3.1 机械案例 变频恒压供水系统的研究与应用	154
6.3.2 生工案例	159
6.3.3 化工案例 纳米二氧化硅气相燃烧制备技术及设备	165
6.3.4 信息案例 利用数据挖掘技术研究高校大学生的文娱生活	170
6.4 理学案例(化学)	177
6.4.1 氟虫酰胺的合成路线研究	177
6.4.2 纺织品中对特辛基苯酚乙氧基醚测定方法的研究	187
6.4.3 2-硝基-4-甲基苯胺的合成	192
6.4.4 Cu-Mn 催化剂对 CO 低温催化氧化作用的研究	195
6.5 科研开题的信息检索	200
6.6 研发过程中实用数据的获取	201
6.7 利用 SciFinder 解决检索中存在的实际问题	207
6.8 信息在企业中的应用	212
6.8.1 可行性报告××制药厂克菌灵项目	212
6.8.2 投资案例 中国黑茶深加工项目投资报告	217
6.8.3 融资案例 河北省唐山赛德 2×5 万千瓦燃煤热电项目	223
6.8.4 融资案例 广东省沙角火力发电厂 B 融资项目	225
6.9 综合案例:金属切割用“NAS”安全切割气的研发、创新到市场推广应用	226

6.9.1 研发阶段	227
6.9.2 产品推广过程	229
6.9.3 成熟市场占有率提高阶段	230
附录 常用电子资源	231
作业样例	240
参考文献	251

第 1 章 信息检索概念

1.1 信息检索

1.1.1 信息检索的含义

信息检索作为一门现代技术,它与计算机几乎同时问世,两者的发展相辅相成,关系十分密切。信息检索(Information Retrieval)一词最早出现于凯尔文·莫尔斯(Calvin W. Mooers)在 1950 年国际数学会议上发表的一篇题为《把信息检索看作是时间性的通讯》的论文中。在此论文中,凯尔文提出了“信息检索是一种时间性的通讯形式”的观点。此后,随着信息环境的变化、用户需求的发展、信息技术的进步,信息检索概念的内涵也几次经历发展变化。目前为公众所普遍接受的信息检索概念,可以从广义和狭义两个方面来讲。从广义上讲,信息检索包括两个过程,一是信息的存储(Information Storage),即信息的标引、加工和存储过程;二是信息的检索(Information Retrieval),即信息用户的查找过程。从狭义上讲,信息检索仅指后一过程。对于信息用户来说,信息检索即如何找出自己所需信息的过程,在此过程中,用户只需要知道如何能够快捷、方便、高效地获取所需的信息,而不必掌握信息的组织管理模式以及信息的存储地点。

信息检索的本质是一个匹配的过程,即信息用户的需求和一定的信息集合进行比较和选择的过程,换言之,也就是用户根据自己的需求提出的主题概念或提问表达式与一定信息系统的检索语言相适应的过程。如果两者相适应并取得一致,则所需信息就被检中,否则检索失败。

现代的信息检索至少包括三层含义:一是按某一主题或某一特征从信息源或数据中查找到相关的信息及其获取线索,如 OPAC 检索;二是为解决某一问题从信息源或数据库中获取隐含于文献中的事实、数据、图像或理论等未知的知识,如全文检索;三是将查询的信息及其知识进行分析加工,提供给用户使用,如科技查新。

1.1.2 信息检索的意义

近代和现代科学史的大量事实证明:没有科学上的继承和借鉴,就没有提高;没有科学工作上的交流和综合,就没有发展。伟大的经典力学的奠基人、英国著名科学家牛顿曾说过:“如果说我比别人看得更远些,那是因为我站在了巨人的肩上。”的确,牛顿就是在意大利的伽利略和德国的开普勒所认识、总结的力学定律的基础上,继承、借鉴、综合和发展了他们的学说,从而达到了经典力学的高峰。

无论是国家、企业,还是个人,在解决问题、分析问题,以及最终做决策时都必须进行信息检索。信息检索的意义主要表现在以下六个方面。

1. 信息控制的手段

现代社会中,信息量呈爆炸性增长。国外统计资料表明:科技成果每增加 1 倍,信息量就

增加几倍;生产翻一番,信息量就增加4倍。据联合国教科文组织统计,全世界发行的图书在20世纪50年代为20多万种,20世纪60年代为40万种,20世纪70年代为60万种,20世纪80年代为80万种,平均每20年图书品种就增加1倍,图书册数增加2倍。全世界的期刊在20世纪50年代约为2万种,20世纪60年代约4万种,20世纪70年代约8万种,20世纪80年代约15万种,基本上是每10年翻一番。全世界的专利文献每年增加40万件,技术标准增加20万件以上,产品资料增加50万~60万件,而会议文献每年出版量则高达100万篇。苏联相关报刊曾报道,列宁图书馆的书架展开长度已达400千米,而且以每年15千米的速度延伸;而美国国会图书馆的书架展开长度已超过了858千米。

在当今的大数据时代,信息的迅猛增长与读者对信息的需求之间的矛盾日益尖锐,对信息进行控制的要求也越来越强烈。近十年来,世界上电子数据增长的规模和速度是异常惊人的。美国国家安全局自2001年,也就是“9·11”事件发生前7个月开始,就一直在不分类别地搜集和保存所有美国公民的详细电话通信记录,5年内这一数量就多达1.9万亿条。由于涉嫌侵犯公民的隐私,这一情况直到2006年被《今日美国》披露后受到了民众广泛的质疑。现如今,数据量的增长更是指数级的。2009年,美国政府产生的数据达848 PB,并且仅美国的医疗数据就达150 EB。2011年全球产生的数据量已经达到了1.8 ZB,并且预计每年将翻一番。

信息检索正是为了解决这一矛盾而发展起来的。通过信息检索,可以全面地了解和掌握信息发展的概貌,可以有效地管理各种分散的信息流,也可以在信息的海洋中快速找到所需的资料。

2. 挖掘分析信息,制定高明的市场谋略

信息是知识经济时代的重要资源,善于挖掘和分析信息,特别是竞争情报,往往能制定出高明的市场谋略,从而出奇制胜。竞争情报是关于竞争环境、竞争对手和竞争策略的情报信息,研究,是为了提高企业竞争力而进行的一切正当合法的情报活动,其核心就是知己、知彼、知环境。

在人类发展史上,各种带有“竞争性”的情报活动早就存在于政治、军事、经济、技术等领域。随着经济全球化趋势的加快以及信息技术的迅速发展,企业间的竞争越来越激烈,信息资源在竞争中的作用日趋重要,企业信息化已成为企业实力与地位的重要标志和象征。竞争情报促进了企业信息化的发展,企业要在竞争上赢得优势,必须依靠竞争情报不断更新企业面貌,增强企业竞争能力与盈利能力,适应信息时代的发展和激烈竞争的要求。一个不知道发挥竞争情报作用的企业,犹如“盲人骑瞎马,夜半临深池”,终将被激烈的市场竞争所淘汰。竞争情报已经继资本、技术、人才之后,成为企业的第四大核心竞争力,能在很大程度上决定着企业的生死未来。

因此,信息收集、信息组织、信息分析与决策的能力已经成为当代企业领先于竞争者的基本能力,竞争情报知识的学习和竞争情报能力的培养正在成为企业管理新的热点。

1) 竞争情报的作用

竞争情报在企业中的作用主要体现在以下几个方面。

(1) 预警。竞争情报最重要的功能之一是使企业避免受到突然袭击。在市场上,最糟糕的情景就是面对竞争对手的突然行动和突然的市场威胁而自己却毫无准备。竞争情报有助于发现市场上的危险和机会,并通过减少竞争对手的反应时间和增加自己的反应时间来获得竞争优势。一个有效的竞争情报系统可以在监测商业环境、跟踪技术变化、监测主要客户动向、发现潜在竞争对手等方面发挥重要的预警作用。

(2) 决策支持。竞争情报对高层管理人员在企业并购、投资、竞争等方面的战略决策具有积极的作用。每一个决策都不是领导拍脑袋就能解决的,而是需要一定的情报分析支持,当今日趋全球化的市场竞争更需要系统、全面的情报支持。利用竞争情报可以使企业主管增加决策的成功率,在生产决策、竞争方式决策、开发新市场和技术开发等方面做出更科学、合理的决策。

(3) 学习工具。竞争情报能让你了解对手的情况,从而能帮助你不断接触到新的思想、新的理念,因此,竞争对手就是你最好的老师,为你提供经验教训、参照标准。通过对竞争对手的经营、销售和管理等信息的系统收集和分析,可以使你得到技术上的借鉴、进行标杆比较,同时开拓你的管理思路。

2) 竞争情报的常用分析方法

常用的竞争情报分析方法有关键成功因素法、SWOT 法、定标比超法、价值链分析法、情景分析法等。

(1) 关键成功因素(Key Success Factors 或 Critical Success Factors,简称 KSF)法

KSF 法也称短板方法,是为达到目标而必须正确进行的事项。一个完整的 KSF 分析方法主要有 5 个步骤。

- ① 定位;
- ② 识别 KSF;
- ③ 收集 KSF 情报;
- ④ 比较评估 KSF;
- ⑤ 制订行动计划。

(2) SWOT 法

SWOT 法最早是由美国旧金山大学的管理学教授韦里克(H. Weihrich)在 20 世纪 80 年代初提出来的。

SWOT 是英文单词“Strengths”、“Weakness”、“Opportunities”和“Threats”的首字母缩写,是将组织或个人内外各方面条件进行综合和概括,分析优劣势、面临的机会和威胁,在此基础上将内部的资源因素、外部因素造成的机会和风险进行合理、有效的匹配,从而制定良好的战略,以掌握外部机会,规避威胁的分析方法。

(3) 定标比超法

定标比超法是由英文“Benchmarking”翻译而来的,也称为基准调查、基准管理、标高超越、立杆比超等,是将本企业经营的各方面状况和环节与竞争对手或个人定位进行对照分析,通过评价自身和研究其他组织或个人,将外部的优秀业绩作为自身的内部发展目标,并把外界最佳做法移植到自身经营或成长环节中,以提高自身竞争力的一种方法。

(4) 价值链分析法

价值链是企业对产品进行设计、生产、营销、交货等基本活动以及相关辅助活动的集合。价值链分析法将价值而不是成本作为其分析的基础。它将一个企业或个人的活动分为在技术上和物质上界限分明的许多活动,这些活动被称为价值活动。

价值链分析可以分以下三步来进行。

- ① 关注链条整体,确定重点环节;
- ② 抓住重点环节,进行重点分析;
- ③ 环节情况汇总,整体把握链条。

(5) 情景分析法

情景分析法也称前景描述法、情景描述法等,是决胜于未来的一种竞争情报分析方法。它是通过想象、联想、推测和猜想来构思和描绘可能的未来面貌的一种创造性思维活动。情景分析法的主要步骤如下。

① 确定预测的主题、明确预测的范围;

② 建立影响因素数据库,并将影响因素按其影响方向划分成一些恰当的集合;

③ 根据影响因素集合,构思各种可能的未来图景;

④ 设想一些突发事件,预测其对未来情景可能的影响;

⑤ 描述到达未来各种状态的发展演变途径。

(6) 专利分析法

专利分析法是企业竞争情报常用的分析方法之一。常用的通过专利分析技术走势的方法主要是定性分析方法,如加州理工学院的 Jay E. Paap 教授倡导的技术机会分析(Technology Opportunity Analysis,简称 TOA)法。TOA 法主要包括以下四个步骤。

① 搜索。查找相关的数据源并检索出相关领域的电子形式数据。

② 计数。利用软件编辑、计算、排列结果。例如,计算全部高频词、文摘数量、作者及其所属单位,可表明某个主题正在升温还是开始降温,与之相关的活动在时间上呈现何种分布。

③ 联系。采用 TOA 或其他专利分析软件提供的矩阵操作来帮助识别联系,进行聚类,追踪一段时间内的变化,注视正在出现的新事物,以及识别谁是某个领域最活跃的人。

④ 表示。把结果转换到 Excel 或者其他类型的电子表格和图形包里,用图形表示正在发生什么,对检索结果按时间片段显示事件地图是如何随时间演变的。

(7) 标准文献分析法

标准文献是一种特别的文献,几乎所有的标准都有着很强的法规时效性,生产中需要随时关注最新适用的标准,对于查找包括标准文献在内的求证型文献检索课题,一定要尽量保证查全率,以免造成损失。

【案例 1】日本是一个信息大国,非常善于利用信息。20 世纪 60 年代的中国对外严密封锁大庆油田的信息,地图上并没有大庆,因此日本人无法获得大庆油田的信息,于是开始研究所有关于大庆的零星情报。仅靠当时中国的《人民日报》和《人民画报》的几段报道和几幅图片,日本人就获得了深层次的信息,分析步骤如下。

(1) 确定大庆的大概位置。画报封面上有王进喜同志的照片:王进喜身穿大棉袄,天空下着鹅毛大雪;还有一幅照片的背景是广袤的黑色土壤的平原。日本人据此分析大庆应该在东北,否则不会下这么大的雪,土壤也不会是黑色的,但还是无法判断大庆的具体地理位置。《人民日报》有一篇报道提到,王进喜同志到了马家窑,说了一句:“好大的油海啊!我们要把中国石油落后的帽子扔到太平洋里去。”日本人据此分析,马家窑就是大庆的中心;“好大的油海”暗示油田的储油量很大。

(2) 确定大庆的中心位置。《人民日报》报道,中国工人阶级发扬了一不怕苦,二不怕死的革命精神,大庆的设备不用马拉车推,完全是肩扛人抬。日本人分析:油田不会远离火车站,远了就拉不动设备。后来日本方派了一名经济间谍以游客的身份前往东北研究铁路线,结果发现了油罐车,顺着铁路找到了离马家窑不远的车站。

(3) 推算产油量和石油需求量。1966 年,王进喜同志光荣地参加了全国人民代表大会。

根据这一消息,日本人说大庆一定出油了,假如不出油,王进喜当不了人大代表。根据《人民日报》上一幅钻塔的照片,以钻台上手柄的架势与几位操作工人手臂的对比,算出油井的直径是10多厘米,根据直径推算每天的产油量。结合中国国务院的政府工作报告公布的数据,把该年产油量减去上一年产油量,就得出大庆的石油产量,算得很准。日本还预测,大约在1980年以后,中国的工业将有较大发展,会需要大量石油。

(4) 提前设计适合中国专用的石油设备。日本人根据当时中国的工业水平预测,中国将无法自行设计制造石油设备,必然会向全世界公开招标,于是决定根据中国的需求特点预先设计石油化工设备。果然,大庆油田出油后,我国向世界很多国家征求设计都不成功,此时日本人说:“早就为中国设计好了!”结果日本人从中获利不少。

忽略信息分析,有时会痛失市场。

【案例2】日本凭一纸公文开发中国市场。20世纪80年代,北京市政府曾经签发了一份红头文件,内容很简单,即要求政府机关购买小轿车需要限定指标,但是不限定客货两用车的购买指标。这份公文下发到各部门,并没有保密。日本三菱公司驻北京办事处的一位信息员分析这份文件,认为如果根据中国当时较差的公路状况生产一种客货两用的面包车,向中国政府机关推销,就能占领中国市场。于是日本三菱公司抢先生产这种车,很快风靡中国。可是我国的各汽车厂并没有悟出那份公文的市场价值,从而只能痛失市场。

【案例3】反倾销是欧美国家在对外贸易中擅用的兵器之一。而中国企业经常成为它们的目标,目前中国是世界上出口产品受到反倾销调查最多的国家。欧盟早在1998年就制定了CR法规草案,矛头直指浙江省打火机厂商。CR法规即儿童防护法规,该法规规定,售价在2欧元以下的打火机必须安装“安全锁”,并将“使用寿命达5年以上”作为安全锁设置的衡量标准。这一规定,浙江企业无一有能力做到,这意味着,如果该法规被付诸实施,浙江省绝大多数生产打火机的企业都将从欧盟市场“失业”,损失巨大到自己根本无法估算。

3. 有利于为决策提供科学依据

虽然信息本身不能确保决策正确无误,但它是决策的基础。一个国家、地区或组织要发展什么,限制什么,引进什么,都需要有准确、可靠和及时的科技信息作依据,才能做出正确的决策。如果决策者重视和善于利用科技信息,就有可能避免重大损失,还有可能快人一招,从中获益。事实证明,不仅科研人员需要科技信息,计划、管理、决策部门也同样需要科技信息。

【案例4】20世纪70年代末,荷兰飞利浦公司推出数码激光唱片,这项突破性的音响技术吸引了众多欧美大公司纷纷投入巨资设厂生产。日本公司在得知这条信息后,经过细致地研究分析,做出了不放弃原已占领的磁带市场的决策。它们悄悄地研制出效果更佳、功能更强的数码录音带及配套设备,使有些激光唱片公司刚刚投产或刚完成庞大的建设工程便面临严峻的挑战。

【案例5】2008年初,一场罕见的雪灾袭击了我国南方大部分地区,交通运输部的有关领导急需国外雪灾道路应对方案。我国情报人员在掌握国内外大量资料的基础上,采用主题词自动标引技术,迅速标出各资料主题,据此选取相关资料,同时对其进行主题对比与分析,找出了各国在不同环境下的应对措施及其依据和特点,形成了国外雪灾公路应对分析报告,为交通运输部的领导做出决策提供了依据。

当今是“大数据”时代,在商业、经济及其他领域中,决策将日益基于数据和分析做出,而非基于经验和直觉。通过对过去和现在的数据进行分析,它能够精确预测未来;通过对组织内部和外部的数据整合,它能够洞察事物之间的相关关系;通过对海量数据的挖掘,它能够代替

人脑,承担起社会管理的职责。

最早提出“大数据”时代到来的是全球知名的咨询公司——麦肯锡,它认为,“数据已经渗透到当今每一个行业和业务职能领域,成为重要的生产因素。人们对于海量数据的挖掘和运用,预示着新一波生产率增长和消费者盈余浪潮的到来。”

2007年,诺贝尔奖获得者吉姆·格雷提到,数据密集型科学正在从计算科学中分离出来,成为科学研究的第四范式。而此时,一切跨国企业已经关注到数据密集型科学的到来。例如,微软研究院出版了《第四范式:数据密集型科学发现》一书,并延伸出“用海量数据重新定义生态科学”、“让我们更接近太空:海量数据中的发现”、“地球科学的研究工具:下一代传感器网络和环境科学”等相关科研课题。

像经典力学、量子力学和计算科学一样,数据密集型科学必将影响到社会科学的研究方式。《大数据时代:生活、工作与思维的大变革》提出了相关关系的大数据思维,即人们可以驾驭所有数据,而不是仅仅抽取小样本;人们可以挖掘更具混杂性的数据,而不用苛求数据的精准性;人们只需知道“知其然”的相关关系,而不需深究“知其所以然”的因果关系。

【案例6】2012年秋天,全球广告巨头WPP集团首席执行官苏铭天爵士约访谷歌首席执行官拉里·佩奇时,佩奇派出无人驾驶汽车去接他。这是一辆集大量高科技设备于一身、自动驾驶的雷克萨斯SUV汽车,车上配有雷达、传感器和每秒可测量超过150万次的激光扫描仪。它自助行进了20分钟,驶过280号州际公路和繁忙的85号州际公路。汽车通过自动驾驶仪进行自动巡航,自行校正路线,在前方出现车辆和行人时减速,然后又加速驶出临近车辆的盲区,最后来到距谷歌公司约32千米外的瑰丽酒店。

这种无人驾驶汽车是谷歌正在研发的一种未来交通工具。它不仅借助对路况历史和实时的计算智能地制动、并线和超车,而且做到了节约能源(因为它使交通的运行更畅通)、提高生产力(可以减少数小时的通勤时间而用于其他事务)。在到达目的地之后,汽车甚至可以自动驶向通过大数据导航的停车位。如果你还需要出行,只需借助手机等移动终端就能指挥汽车到达预定地点。

科研范式的转变最终反馈到人们的思维模式和决策模式的转变。谷歌的无人驾驶汽车就是基于大数据的分析,借助计算技术和人工智能实现了交通引导和控制功能。通过无处不在的计算和传感器,大数据能够解析存在于现实世界、虚拟世界以及虚实融合世界的复杂网络关系,并适时做出判断和决策。这种决策模式遵循数据转变为信息、信息转变为知识、知识涌现出智慧的流程。

大数据至关重要的方面,就是它会直接影响组织怎样做决策、谁来做决策。在信息有限、获取成本昂贵且没有被数字化的时代,组织内做重大决策的人,都是典型的位高权重的人,要不然就是花高价请来的拥有专业技能和显赫履历的外部智囊。但是,在今时今日的商业世界中,高管的决策仍然更多地依赖个人经验和直觉,而不是基于数据。

大数据推动下的管理变革将力求改变这一现状,让数据做主。美国电影《点球成金》非常直观地显示了这一变动趋势。奥克兰棒球队白象队总经理比利·比恩抛弃了几百年一直依赖的选择球员的传统惯例,采用了一种依靠电脑程序和数学建模分析比赛数据来选择球员的方法。最终,他带领球队在2002年美国联盟赛中夺得冠军,还取得了20场连胜的战绩。从那以后,统计学家取代了球探成为棒球专家,而联盟的其他球队也开始争先采用白象队的方法来指导球队。

人类依靠自身判断做决策到依靠数据做决策的转变,是大数据的最大贡献。行业专家和

技术专家的光芒将因为统计学家和数据分析专家的出现而黯然失色,因为后者不受旧观念的影响,能够聆听数据发出来的声音。行业专家不会消失,但是他们必须与数据表达的信息进行博弈。直觉的判断被迫让位于精准的数据分析,这将迫使人们调整在管理、决策、人力资源和教育方面的传统理念。

如今,我们正在见证专家在各个领域影响的减弱。在传媒界,《赫芬顿邮报》和 Gawker 网站上传播的新闻通常取决于数据,而不再取决于编辑的新闻敏感度,数据比有经验的记者更能揭示出哪些是符合大众口味的新闻。实际上,Digg 网站完全不自主提供新闻,也没有编辑,它是一个文章投票评论站点,用户可以随意提交文章,当其他用户认为文章不错时,dig 一下。当 dig 数达到一定程度时,该文章就会出现在网站首页,网站将决策权完全交给数据。到 2006 年,Digg 网站的 Alexa 全球排名达 100 位,轻松打败福克斯新闻网,逼近《纽约时报》。

大数据所具有的在区域之间、行业之间和企业部门之间的穿透性,正在颠覆传统的、线性的、自上而下的精英决策模型,正在形成非线性的、面向不确定性的、自下而上的决策基础。正如美国管理与统计学家爱德华·戴明所说:“我们信奉上帝。除了上帝,任何人都必须用数据说话。”

4. 有利于减少课题的重复研究、提高科研成功率

任何科学研究都是在继承前人的知识后有所发明、有所创新的。也就是说,每个人都把前人认识事物的终点作为继承探索的起点。任何人从事某一特定领域的学术活动,或开始做一项新的科研工作,都要花费大量的时间,对有关文献进行全面的调查研究,摸清国内外是否有人做过或者正在做同样的工作,取得了一些什么成果,尚存在什么问题,以便借鉴、改进和部署自己的工作。只有这样方能做到心中有数,才能有所发现、有所创新、有所前进,否则容易造成重复劳动,导致人力、物力、财力的浪费。

文献资料反映了国内外研究的学术思想和最新成就,是我们了解科研前沿动向并获得新情报、新信息的有效途径。通过查阅文献资料,了解国内外最新的理论、手段和研究方法,从过去和现在的有关研究成果中受到启发,使研究范围内的概念、理论具体化,而且为更科学地论证自己的观点提供有说服力的、丰富的事实和数据资料,使研究结论建立在可靠的材料基础上。

【案例 7】日本高能物理研究所,由于检索和利用国外的情报资料,研制成功的第一台高能加速器的投资为 40 亿美元,仅为国外同类产品投资费用的 50%。

【案例 8】我国葛洲坝工程二江电站出线方案,由于情报人员及时搜集、查阅、分析了大量国内外情报资料,提出了高压架线路方案,该方案被采纳后,仅投资一项就节约了 400 万元。

【案例 9】汉字激光照排系统的发明人王选教授,走的正是这样一条“捷径”。1986 年,他只是北京大学一名助教,仅有 10 万元科研经费,要研制取代铅字印刷的新技术。当时国内权威都认为应该跟着日本人的步伐,完善光学机械式印刷系统。但是王选就是不肯盲从权威,开题立项之前曾用了一年的时间,检索和研究了大量国外专利信息,了解到照排技术从“手动式”、“光学机械式”、“阴极射线管”已经发展到第四代即“激光照排”,但是激光照排还不完善,国外尚无商品。于是,王选便越过当时日本流行的光机式、欧美流行的阴极射线管式,直接研制成功第四代激光照排系统,实现了跨越式发展,节约了科研经费和时间。

相反,如果在科学研究中,忽视信息检索,不能做好继承和借鉴工作,则容易重复研究,浪费大量人力、物力和财力,或者采用了较差的实验方法,使研究工作进展缓慢甚至失败。据统计,20 世纪 80 年代,各国因未查阅专利信息,使研究课题失去研究价值,每年造成的损失如美

国达 11 亿~13 亿美元,英国为 2 000 万~3 000 万英镑,苏联为 16 亿~19 亿卢布。欧洲专利局的一项研究结果表明:十几个欧洲专利条约成员国在应用技术的研究开发中,由于利用了专利信息,避免了重复研究,每年可节约 300 亿马克的研究开发经费。

【案例 10】上海市一家保温瓶生产厂,自力更生进行科技攻关,花了多年时间,据说投入上百万元经费,解决了以镁代银镀膜工艺,20 世纪 80 年代准备申请国家发明奖,但是经上海市科技情报所检索信息,才发现英国的一家公司早在 1929 年就为此项工艺申请了专利。该专利的法定保护期早已期满,成为公知、公用的技术,直接使用就可以了,不需要支付任何费用,而且这家保温瓶生产厂旁边的科技情报所就收藏了该专利说明书。

5. 有助于节约时间、提高科研效率

随着科学技术的发展,文献数量在急剧增加并且学科间相互渗透,科研人员在进行一项科研活动中,查找资料就花费了大量时间。据美国和日本在 20 世纪 60 年代的一项统计结果表明,科学工作者在从事科研活动所花的时间为:试验研究占 32.1%,计划、思考占 7.7%,数据处理占 9.3%,查找情报资料占 50.9%,如果熟悉文献检索方法,就能大大节省查找资料的时间,从而加快科研速度,早出科研成果。

【案例 11】“汪克尔”转子发动机是德国人发明的,日本开展此项研究的时间比德国晚 10 年,但由于日本政府和有关公司全力以赴地搜集、检索德国这方面的情报,结果使得日本先于德国 10 年将装有这种发动机的小汽车投入市场,前后共计为日本赢得了 20 年的时间。

【案例 12】北京低压电器厂开发漏电保护开关,经过一年多的探索仍没有找到理想的技术方案,后来利用 7 天的时间查阅了大量的专利信息,找出 70 多篇相关技术资料,经过对专利信息的分析,大受启发,仅用了 3 天就制定出一套切实可行的技术方案,研究开发出具有国际先进水平的新产品,并申请了专利。

在当今世界,提高科研效率,加快科研速度的意义还在于使相同科研课题在国内外竞争中处于有利位置。国际专利法规定,对相同的发明成果,按优先申请原则授予专利权,即只授予第一个申请人专利权,其后申请的发明作为已知技术处理。显然,如果忽视科研速度,即使科研获得了成功,但由于发明失去了时间的新颖性,也会变成无效劳动,给国家带来损失。

6. 掌握获取信息的方法,提高信息意识和信息素养

无论从事哪项研究工作或对某一问题进行研究探讨,或对某一技术、工艺进行改革,都必须从查找文献、寻求信息开始,研究成功的可能性才越大。作为一个现代科技工作者,不仅要具有实际的科学研究能力,而且还应具有文献收集、选择和利用的能力。因此,学习信息检索,不仅要学会查找文献的方法,更重要的是要通过本课程的学习,提高自身的信息意识和信息素养,提高独立分析问题和解决问题的能力,使自己具有更强的社会生存和社会竞争能力。

1) 信息意识

信息意识属于意识形态范畴,是意识的一种,是信息在人脑中的集中反映。具体而言,信息意识就是人作为信息的主体在信息活动中产生的知识、观点和理论的总和。它包括两方面的含义:一方面是指信息主体对信息的认识过程,也就是人对自身信息需要、信息的社会价值、个人活动与信息的关系及社会信息环境等方面的自觉心理反应;另一方面是指信息主体对信息的评价过程,它包括对待信息的态度,对信息质量的变化等所做出的评估,并能以此指导个人的信息行为。

信息意识从信息认知、信息情感和信息行为倾向这三个层面体现出来。信息认知是指对信息和信息活动的了解和看法,其中最重要的是评价性的认知;信息情感是指人们在感受信息

的过程中,逐渐形成的反映需求关系的内心体验,这种体验相对持久、稳定,而不是即时产生的情绪;信息行为倾向是指个人在信息活动中表现出来的行为趋势,是信息行为的心理准备状态。人们的信息收集活动是受信息需求驱使的,而影响需求的力量大小主要就是需求被意识的清晰程度,即意识越明确,行动目标越清楚,则信息活动的动机就越稳定、持久、强烈,努力程度也就越高。如果这种意识仅停留在感性阶段,那么接收信息总是处于被动状态;反之,信息意识经常在觉醒、活跃的状态,就会促使人们主动制订信息活动计划,预见到各种变化,并做出积极的选择。

能否意识到何时需要信息和需要什么样的信息,是信息意识强弱的最重要表现。强烈的信息意识表现为对信息有敏锐的感受力、洞察力和持久的注意力,以及对信息价值准确的判断力。在具体实践中,信息意识又可分为信息主体意识、信息获取意识、信息传播意识、信息守法意识和信息更新意识等。

信息意识的强弱决定着人们捕捉、判断和利用信息的自觉程度,影响着人们利用信息的能力和效果;同时,它也直接关系到信息素养的整体培养。信息素养是一系列的能力,而不单单是某一种技能,它要求个人不仅能有效地使用所需信息,而且能在需要信息的时候有意识地加以识别和评价。美国图书馆协会和美国教育传播与技术协会于1989年提交的《关于信息素养的总结报告》中,也曾明确地指出:一个有信息素养的人必须能够认识到何时需要信息,需要什么样的信息;评价和使用所需信息,以达到有效利用的目的。由此可见,信息意识是整个信息素养的前提,是信息素养培养过程中关键的一环。它直接影响着信息主体在信息活动中的效能,成为人们从事信息活动的精神驱动力。

2) 信息素养

“信息素养”(Information Literacy)这一概念最初是从西方图书馆检索技术发展和演变而来的,并迅速从图书情报界扩展到教育界乃至社会各界。自从1974年车可斯基提出“信息素养”这一概念至今,很多专家、学者和学术机构研究信息时代的新变化,不断地进行理论与实践的探索,根据自己的研究对信息素养给出了不同的定义,并提出了信息素养的标准。其中被引用较多、影响较大的定义大部分来自美国。国外从1974年到2000年共有18个涉及信息素养的定义,国内关于信息素养的定义有5个,均是在2000年以后界定的。下面列出几个有代表性的定义。

(1) 1983年,美国信息学家霍顿(Horton)提出,教育部门应开设信息素养课程,以提高人们对电子邮件、数据分析以及图书馆网络的使用能力。实际上,这已经把信息素养与计算机和网络联系了起来。

(2) 1989年,美国图书馆协会下属的“信息素养总统委员会”将信息素养定义为“要成为一个有信息素养的人,必须能够确定何时需要信息,并已具有检索、评估和有效使用所需信息的能力”。从此,信息这个概念跨越了图书情报界,迅速扩展到教育界甚至全球各个领域。

(3) 1992年,道伊尔(C. S. Doyle)在《信息素养全美论坛的总结报告》中进一步将信息素养定义为:一个具有信息素养的人,他能够认识到精确和完整的信息是做出合理决策的基础,确定对信息的需求,形成基于信息需求的问题,确定潜在的信息源,制定成功的检索方案,从包括基于计算机的和其他信息源中获取信息、评价信息、组织信息并实际应用,将新信息与原有的知识体系进行融合,在运用批判的观点思考和解决问题的过程中使用信息。这个定义使信息素养的内涵得以具体化。

(4) 1998年,在全美图书馆协会和美国教育传播与技术协会的出版物《信息能力:创建学