

<<数据学>>

图书基本信息

书名：<<数据学>>

13位ISBN编号：9787309069563

10位ISBN编号：7309069560

出版时间：2009-12

出版时间：复旦大学出版社

作者：朱扬勇

页数：131

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## 前言

科学研究所用的方法是逻辑推理和实验，逻辑推理依靠数学，实验依靠观测。

在计算机出现后，科学研究开始使用计算机技术，使得科学研究增加了计算的方法。

在自然科学的研究过程中，遇到了大量的计算问题，这些计算是手工无法完成的，因此计算机的大规模计算能力在科学计算方面获得了很好的应用，并逐渐成为一种科学研究的新方法，相继出现了计算数学、计算物理、计算化学、计算生物学，等等，并且计算生物学已经成为现代生物学的核心方法之一。

后来，科学研究的对象也信息化了，变成了计算机中的数据，最典型的是生命科学领域中基因的信息化，形成由ACGT这4个字母组成的DNA序列数据，研究对象变成了DNA序列数据，出现了生物信息学。

同样的状况也出现在其他研究领域，于是有了脑信息学、地理信息学、行为信息学、社会信息学、经济信息学、历史信息学，等等。

随着国民经济和社会的信息化进程，自然界中的事物以数据的形式存储到计算机系统中，即信息化是一个生产数据的过程。

这些数据是自然和生命的一种表示形式，这记录了人类的行为，包括工作、生活和社会发展。

数据被快速大量地生产并存储到计算机系统中，这种现象称为数据爆炸，数据爆炸在计算机系统中形成数据自然界。

目前，数据爆炸还在进行中，人类还不能清晰地描述数据自然界。

## <<数据学>>

### 内容概要

《数据学》介绍了数据学的由来、基本概念和基本原理，包括：数据大爆炸、数据自然界、数据学基础；介绍了数据学的主要方法，包括：数据勘探、数据获取与整合、数据挖掘、数据实验；还介绍了数据学的应用和数据学面临的挑战。

《数据学》主要作为科学研究工作者的参考书，试图向科学工作者展示一种新的科学，并且能够利用这种新的科学为当前的科学研究工作服务。

《数据学》基本以较为通俗化的语言来表达数据学的基本原理、方法和技术，希望对各领域的科学家，包括行为科学家和社会科学家能够有所启迪。

## 作者简介

朱扬勇，1963年生，浙江武义人。

复旦大学计算机科学技术学院教授，数据科学研究中心主任；上海市政府信息化专家，上海市计算机学会、生物信息学会理事等。

主持完成国家自然科学基金、“863计划”等10多项科研课题，发表论文100余篇，出版图书7本。

目前作为主要倡导者，从事数据科学的研究工作。

熊赅，1980年生，江西南昌人。

复旦大学计算机科学技术学院教师。

主要研究领域为数据科学，发表论文20余篇。

作为项目负责人，主持国家自然科学基金、上海市科委重点发展基金项目等的研究工作，并参与国家“863计划”、“973计划”等多项研究课题。

## 书籍目录

第1章 绪论1.1 数据1.1.1 数据的概念1.1.2 数据与物质1.2 数据爆炸1.3 数据自然界1.3.1 数据不为人类所控制 1.3.2 数据的未知性1.3.3 数据的多样性和复杂性1.4 数据学1.4.1 为什么需要数据学 1.4.2 数据学的概念1.4.3 数据学的框架1.5 与其他科学的关系1.6 小结第2章 数据自然界基础知识2.1 数据自然界的发展2.1.1 3个阶段2.1.2 数据集2.2 面临的问题2.3 数据簇2.3.1 数据的属性2.3.2 相似性与相似函数2.3.3 数据簇2.4 数据分类学2.4.1 数据本体2.4.2 数据分类学2.5 小结第3章 数据勘探3.1 为什么要做数据勘探3.2 什么是数据勘探3.2.1 数据勘探做什么 3.2.2 数据勘探步骤3.2.3 数据矿床3.3 勘探数据集的总体特性3.3.1 通过样本分析判断数据特征3.3.2 如何抽样3.3.3 通过查询判断数据集的特征3.4 勘探数据集的结构3.5 数据工具的勘探3.6 小结第4章 数据获取与整合4.1 数据源存在的问题4.2 数据获取4.2.1 数据获取的方法4.2.2 数据质量4.2.3 数据清洁4.3 数据整合4.3.1 数据整合的动因 4.3.2 数据整合的概念4.3.3 数据整合的主要工作4.3.4 数据整合的方式4.4 数据仓库4.4.1 数据库的局限4.4.2 基本概念4.4.3 数据组织4.5 小结第5章 数据挖掘5.1 数据挖掘的故事5.2 什么是数据挖掘5.2.1 数据挖掘的定义5.2.2 数据挖掘的过程5.3 数据挖掘的任务5.4 数据挖掘的类型5.4.1 一般数据源的挖掘5.4.2 特殊应用数据源的挖掘5.5 小结第6章 数据实验6.1 数据观察6.2 数据实验及其目的6.2.1 什么是数据实验6.2.2 数据实验的目的6.3 数据实验的步骤6.4 小结第7章 数据学应用7.1 科学研究信息化7.2 生物信息学7.2.1 生物数据管理与整合7.2.2 生物数据分析7.3 脑信息学7.3.1 脑信息学研究方法7.3.2 脑数据管理与整合7.4 其他信息学7.5 小结第8章 面临的挑战8.1 数据学理论体系8.1.1 观察与猜想8.1.2 数据运算8.2 数据自然界与人8.2.1 在数据自然界中生活8.2.2 数据搜索8.2.3 数据真实性8.3 数据资源的保护与开发8.3.1 数据资源 8.3.2 数据资源的保护8.3.3 数据资源的开发8.4 小结参考文献

## 章节摘录

插图：信息化的本质是将现实世界中的事物以数据的形式存储到计算机系统中，即信息化是一个生产数据的过程。

这些数据是自然和生命的一种表示形式，这些数据还记录了人类的行为，包括工作、生活和社会发展。

今天，数据被快速大量地生产并存储在计算机系统中，这种现象称为数据爆炸（data explosion）。

数据爆炸在计算机系统中形成数据自然界（data nature）。

研究数据自然界是研究自然界（real nature）的一种有效方法，例如：可以通过研究数据来研究生命（生物信息学）、研究人类行为（行为信息学）。

数据学（dataology）或数据科学（data science）是探索数据自然界奥秘的理论、方法和技术。

本章介绍了数据爆炸、数据自然界和数据学的基本概念，并给出了数据学的定义及其基本框架。

计算机系统中存放的是数据，“数据”的含义很广，不仅指工011、8084.这样一些数字，还指“dataology”、“小舟扬帆出海”、“11/11/11”等符号、字符、日期形式的数据。

确切地说，本书讨论的数据是指能够输入到计算机中的任何东西，如：数字、字符、声音、图像、照片，等等，并且处理数据的计算机程序本身也是“数据”。

<<数据学>>

编辑推荐

《数据学》主要内容简介：复旦大学出版社

#### 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>