

□ 教育部人文社会科学研究项目基金资助 (09YJC870016)

云图书馆

理论与实践

Theory and Practice of Cloud Library

◎王红 著

山西出版传媒集团

山西人民出版社

013029834

G250.7

37

◎王红 著



Theory
and Practice
of
Cloud
Library



图书馆 理论与实践

G250.7
37



北航

C1635532

山西出版传媒集团

山西人民出版社

图书在版编目 (CIP) 数据

云图书馆理论与实践 / 王红著. —太原: 山西人民出版社, 2012. 12

ISBN 978-7-203-08072-5

I. ①云… II. ①王… III. ①计算机网络—应用—图书馆工作—研究 IV. ①G250.7

中国版本图书馆CIP数据核字 (2013) 第022829号

云图书馆理论与实践

著 者 王 红
责任编辑 武 静

出 版 者 山西出版传媒集团·山西人民出版社
地 址 太原市建设南路21号
邮 编 030012
发行营销 0351-4922220 4955996 4956039
0351-4922127 (传真) 4956038 (邮购)
E-mail sxskcb@163.com 发行部
sxskcb@126.com 总编室
网 址 www.sxskcb.com

经 销 者 山西出版传媒集团·山西人民出版社
承 印 者 太原市苹果数码印刷有限公司

开 本 787mm×1092mm 1/16
印 张 12.5
字 数 260千字
印 数 1-1000册
版 次 2012年12月第1版
印 次 2012年12月第1次印刷
书 号 ISBN 978-7-203-08072-5
定 价 38.00元

如有印装质量问题请与本社联系调换

写在前面的话

图书馆在利用网络为用户提供方便快捷的服务，提升社会地位的同时，遇到了一些困境和挑战。如何在现有的技术水平情况下，利用现有的条件解决图书馆所面临的上述问题，是当今图书馆在网络环境下所要解决的重要难题，而云计算技术的出现无疑为上述问题的解决提供了一种崭新的、较为理想的方法。

对于图书馆人而言，虽可能不大理解云计算的内涵，但或多或少已经体验过或正在使用云计算服务，重要的是理解云计算实质上是一个有关数据和运行数据计算设备范式如何转变的问题。这些问题与数字化生活始终紧紧相连。而云计算，把重点转到远离这些设备能够有效地存储数据和能够运行的应用和设备上，提供了最简单的获取数据和应用程序，这些存储在不同地方的互联网上。

山西财经大学图书馆王红老师在其著作中着重论述了云计算在图书馆的具体运用之意义，强调指出在图书馆“云”中，用户可以实时获得云平台中的资源，各成员馆的相关电子资源也储存在云服务器中，实现最大限度的开放存取。更重要的是各图书馆不必花大量金钱购买相同的电子资源，电子资源运营商也不必单独开发异构数据库进行编目、借阅和进行文献传递。

OCLC的WorldCat Local（一站式信息挖掘和获得）就是OCLC使用

云计算的技术向图书馆提供的服务。因此,每一所使用WorldCat Local 服务项目的图书馆,不需要购买任何电脑硬件,也不需要开发任何软件,更不需要有馆员管理任何系统运行。这些服务完全由OCLC通过云计算的技术,提供给图书馆! WorldCat Local的最大特点是把本地检索、全球检索、联邦检索和元数据检索完美地集成在一起;云计算,OCLC叫Web-Scale图书馆管理服务,大规模使用万维网作图书馆信息服务,其作用可以取代目前各图书馆使用的图书馆集成系统,主要是通过全球图书馆合作,除去不需要的重复劳动,或不需要的重复投资,提升合作馆的信息服务功能。

我们新时代的图书馆人必须首先树立云的理念,要有“大图书馆、大服务”的意识,进一步促使云计算在图书馆领域的运用,避免传统图书馆集成系统OPAC上的所有弱点,并增强图书馆信息服务,提升图书馆传递全文功能,通过OCLC的WorldCat,使参加这个策略的图书馆成为万维网的一部分,使图书馆的信息彻底网络化和国际化。同时,使参加这个策略份额图书馆信息使用者,能一次检索到全球图书馆的信息。

丘东江

(OCLC北京代表处首席代表)

2012年12月于北京中关村

目 录

第一章 计算和计算机	001
第一节 计 算	001
第二节 计算机运算与人类思维	005
第三节 人工智能	008
第二章 云计算	017
第一节 云计算的起源	018
第二节 云计算的定义	019
第三节 云计算的特点	020
第四节 云计算的应用形式	022
第五节 云计算的结构	023
第六节 云计算的技术	024
第七节 云计算应用实例	026
第八节 云计算厂商	027
第三章 开源软件和图书馆	034
第一节 开源软件简介	034

第二节 开源软件的发展现状	037
第三节 开源软件在图书馆应用的情况	051
第四节 数字图书馆开放源码应用的商业模式	054
第五节 数字图书馆的开源软件商业模式	056
第四章 云图书馆及其平台的架构与实现	061
第一节 计算与图书馆	061
第二节 图书馆云计算	065
第三节 云图书馆平台的架构与实现	075
第五章 物联网和云图书馆	086
第一节 物联网概述	086
第二节 物联网环境下图书馆的发展趋势	087
第三节 基于物联网的智能云图书馆架构与实现	088
第六章 云图书馆安全、标准及法律问题	098
第一节 云图书馆安全问题	098
第二节 云图书馆标准问题	100
第三节 云计算环境下图书馆与著作权保护	101
第七章 云图书馆组织管理	109
第一节 云图书馆的成员构成	109
第二节 云图书馆的组织形态	110
第三节 云图书馆的组织机制	112
第四节 云图书馆的组织特征	114

第五节	云图书馆的组织管理对象	115	目
第六节	云计算时代图书馆转型发展	117	录
第八章	基于云计算的泛在图书馆个性化知识服务	124	
第一节	云计算和泛在图书馆	124	
第二节	基于云计算的泛在图书馆建设的技术	126	
第三节	泛在图书馆个性化知识服务及其云构成	127	
第九章	基于SoLoMo的云图书馆知识服务模式	133	
第一节	SoLoMo简介	133	
第二节	SoLoMo的特征	135	
第三节	SoLoMo与图书馆	136	
第四节	结 语	143	003
第十章	云计算与图书馆大数据	145	
第一节	大数据简介	145	
第二节	图书馆大数据起源与发展阶段	146	
第三节	图书馆大数据的类型	150	
第四节	云计算图书馆的大数据管理与应用	152	
第十一章	云图书馆实践与探索	157	
第一节	整体思路	158	
第二节	财经科技文献资源“云”平台的架构	159	
第三节	构建财经科技文献资源云平台	162	
第四节	支撑技术	167	

第五节 结 语	178
第十二章 OCLC与云图书馆	179
第一节 基于云计算的图书馆设计与实现	179
第二节 OCLC云图书馆	179
第三节 OCLC云计算述评	184
第四节 从OCLC看图书馆云计算的未来	186
第五节 结 语	188
参考文献	189
附 录	196
后 记	199

第一章 计算和计算机

第一节 计 算

一、计算是什么

001

计算，在计算机出现以前主要是数学领域专有的名词，在计算机出现以后，计算的含义被广泛延伸。我们谈论计算是什么，首先就需要谈数学在发展过程中，对人类提供了哪些精神财富。

中国传统数学在人类的生产和应用方面，以算法为中心，注重解决实际问题，属于应用数学，但是在文化和精神领域，又总是被打上中国哲学与古代学术思想的烙印，往往与术数交织在一起。16世纪末西方数学传入，由于其符号化的特征以及严谨的逻辑体系和演绎方法，使得西方数学逐渐成为主导。

西方数学发展到16世纪，牛顿和莱布尼茨在微积分领域取得了一系列成就，为现代数学应用向两个方向发展提供了巨大的思想武器。牛顿用微积分这一有力的数学武器，激励后代的科学家向人类未知的领域，展开了卓有成效的探索。莱布尼茨的数学思想，则为后代的学者用符

号、公式和公理的方法研究人的思维过程、思维规律，提供了有力的指导。莱布尼茨认为通过建立一种精确的、普遍的符号语言，并寻求一种推理演算，能够解决人如何推理的问题。自莱布尼茨之后两个世纪以来，许多逻辑学家、数学家、发明家和工程师沿着莱布尼茨的思路进行了大量实质性的研究和探索。在数理逻辑、数学、机器计算等方面取得丰硕成果。到20世纪40年代，计算机出现的理论和物质基础已经成熟。

二、神秘的二进制

我们知道，现代计算机普遍是用二进制进行运算的，这是因为机器对电路的判断只有接通和断开两种状态，恰好可以用“0和1”来表示。

但是人们根据日常的生活经验，在历史的进化过程中，逐渐形成使用十进制为主流的计数和运算，相对于人的十指，十进制有0, 1, 2……9共10个基本数字，采用“逢十进一”的运算规则，十进制的特点是需要10个数字符号。那么有没有一种包含最少计数符号，也能表示不同数量的技术方法呢？

莱布尼茨1679年发表的《二进位数学》，第一次清晰阐明了二进制的运算规则。在莱布尼茨的二进制中，采用了“逢二进一”的运算规则，通过0与1两个符号，就可以表示一切数字。

1703年莱布尼茨在法国《皇家科学院院刊》上，发表了论文《二进位算术的阐述——关于只用0和1兼论其用处及伏羲氏所用数字的意义》，进一步阐述了二进制的哲学意义。在德国图灵根著名的郭塔王宫图书馆（Schlossbibliothek zu Gotha）保存着一份珍贵的手稿，其标题为：“1与0，一切数字的神奇渊源。这是造物的秘密美妙的典范，因为，一切无非都来自上帝。”他在写给当时在中国传教的法国耶稣士会牧师布维（Joachim Bouvet, 1662 - 1732）的信中说：“只有当我们仅仅用0和1来表达这个数字时，才能理解，为什么第七天才最完美，为什么7是神圣的数字。特别值得注意的是它（第七天）的特征（写作二进制的111）与三位一体的关联。”

二进制的产生与中国古代哲学的源头《伏羲八卦图》有着密切的联系，不但表现出中西方文化通过二进制进行的一次完美交融，因为在易经八卦中，通过阴阳引申，就可以表示宇宙万物及其相生相克的原理；而且表现出现代精密严谨的科学技术与古代人类朴素的认知，赋予人类挑战未知领域不断进取的精神动力。如今电子计算机在经历早期通过纸带进行输入输出的阶段后，发展到现在已经可以模拟和运算世界上的一切事物，小到微观量子级构成物质的基本粒子，大到天文尺度的宇宙演化，都已经成为计算机施展身手广阔的舞台。

三、一切皆可计算

1935年，对数理逻辑发生兴趣的阿兰图灵提出，可以把推理当作进行简单机械动作的计算机运算抽象模型——图灵机模型，图灵的基本思想是用机器来模拟人们用纸笔进行数学运算的过程，为了模拟人的这种运算过程，图灵构造出一台假想的机器，他把人运算的过程看作下列两种简单的动作：

(a) 在纸上写上或擦除某个符号。

(b) 把注意力从纸的一个位置移动到另一个位置。

在每个阶段，人要决定下一步的动作，依赖于 (a) 此人当前所关注的纸上某个位置的符号和 (b) 此人当前思维的状态。

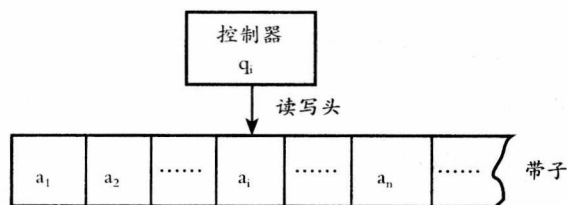


图1 图灵机

图灵机采用以下方式模拟人的运算过程：

(1) 一条无限长的纸带 TAPE。纸带被划分为一个接一个小格子，每个格子上包含一个来自有限字母表的符号，字母表中有一个特殊的符号表示空白。纸带上的格子从左到右依次被编号为 0, 1, 2, …… 纸带的右端可以无限伸展。

(2) 一个读写头 HEAD。该读写头可以在纸带上左右移动，它能读出当前所指的格子上的符号，并能改变当前格子上的符号。

(3) 一套控制规则 TABLE。它根据当前机器所处的状态以及当前读写头所指的格子上的符号来确定读写头下一步的动作，并改变状态寄存器的值，令机器进入一个新的状态。

(4) 一个状态寄存器。它用来保存图灵机当前所处的状态。图灵机的所有可能状态的数目是有限的，并且有一个特殊的状态，称为停机状态。

图灵机执行一步工作的过程：

读写磁头在所扫描的方格上写上符号，原有的符号消除，磁头移动到下一个方格，机器状态转向另一个状态。如此周而复始，直到遇到命令及其停止的状态。

这个看似简单的图灵机，就是现代计算机最基本的理论和结构的原型。事实上，这个图灵机，只要具备足够的时间和空间，即允许运行足够多的步数、无限长的磁带，拥有更快的运行速度，就可以代替目前任何的计算机。

美国著名的数学家邱奇 (Alonzo Church) 指出：任何计算如果存在一个有效过程，它就能被图灵机所实现，这就是有名的邱奇论题，图灵也对图灵机的计算能力充满信心，著名的图灵论题指出：凡是可计算的函数，都可以用图灵机来计算。

第二节 计算机运算与人类思维

一、计算机产生

在人类文明发展历史的长河中，计算工具也经历了从简单到复杂、从低级到高级的发展过程。如曾有“结绳记事”的绳结、算筹、算盘、计算尺、手摇机械计算机、电动机械计算机等。它们在不同的历史时期发挥了各自的作用，而且也孕育了电子计算机设计思想的雏形。

随着计算机的处理能力飞速提高，计算机也开始由二进制代码——自然符号，向与人类思维密切相关的符号——语言代码处理和思维的本质领域——人工智能等方向进军。

二、计算机思维的载体——计算机语言

计算机是由硬件和软件两大系统构成的，硬件是计算机的物质基础，软件则是计算机的灵魂，通过计算机编程语言实现。计算机的编程语言与人类的语言符号系统极其相似，计算机的运算与人类的思考方式也极其近似，了解计算机语言的运行方式，可以在一定程度上了解计算机运算与人类自然思维之间隐藏的深刻关系，为计算机处理人类自然语言符号系统、模仿人类思维提供有益的借鉴。

（一）机器语言及思维方式

计算机语言的发展是一个不断进化的过程，早期的计算机语言是由二进制的“0”、“1”组成的机器直接识别的语言，计算机计算要通过一串串由0和1组成的指令序列，通过打孔的纸带交由计算机执行，执行

运算的结果也是通过对纸带进行打孔输出的。这些由0、1代码组成的程序、输入输出的纸带，对程序员来说很难辨别和记忆，使用机器语言是十分痛苦的差事，特别是在程序有错需要修改时，更是加重了程序人员的劳动。

早期的计算科研人员为了减轻使用机器语言繁杂的劳动和编程的痛苦，在二进制机器语言基础上发明了汇编语言，用一些简洁的英文字母、符号串来替代一个特定指令的二进制串，助记符号（Memoni）代替了操作码，地址符号（Symbol）或标号（Label）代替了地址码，机器语言的二进制码就变成了符号，因此汇编语言亦称为符号语言。汇编语言虽然是一种机器识别语言，但机器并不能直接识别，需要通过一个专门的程序（即汇编程序或编译器），负责将这些符号翻译成二进制数的机器语言供识别和运算。由于汇编语言的符号非常接近于自然语言的要素，这样一来，人们可以很容易读懂并理解程序在干什么，纠错及维护都变得方便了。虽然汇编语言依旧距离人们使用的自然语言和数学语言相距甚远，在计算机程序开发中很少使用，但由于汇编更接近机器语言，能够直接对硬件进行操作，生成的程序与其他的语言相比具有更高的运行速度，占用更小的内存，在一些对于时效性要求很高的程序、许多大型程序的核心模块以及工业控制方面依然有大量的应用。

（二）第一个接近人类思维的编程语言——FORTRAN

FORTRAN是英文“Formula Translator”的缩写，译为“公式翻译器”，它是世界上最早出现的计算机高级程序设计语言，至今仍广泛应用于科学和工程计算领域。美国IBM公司约翰·巴克斯（John Backus）针对汇编语言的缺点着手研究开发FORTRAN语言，并于1954年在纽约正式对外发布。FORTRAN语言的最大特性是接近数学公式的自然描述，在计算机里具有很高的执行效率。它简单易学，语法严谨，可以直接对矩阵和复数进行运算，广泛地应用于数值计算领域，积累了大量高

效而可靠的源程序。FORTRAN发明的年代还没有键盘与屏幕，代码还必须使用打卡机打在打孔卡（punch card）上，但FORTRAN已经包含与自然语言相对应的32种语句，如：赋值语句、IF条件语句、DO循环语句、格式化输入输出语句（FORMAT，READ，READ INPUT TAPE，WRITE，WRITE OUTPUT TAPE，PRINT）等。

FORTRAN出现后，很多专用的大型数值运算计算机针对FORTRAN做了优化，广泛地应用于并行计算和高性能计算领域，FORTRAN90、FORTRAN95、FORTRAN2003的相继推出使FORTRAN语言具备了现代高级编程语言的一些特性。从FORTRAN语言诞生之后，软件越来越多，规模越来越大，而软件的生产基本上是各自为阵，缺乏科学规范的系统规划与测试、评估标准，大批耗费巨资建立起来的软件系统，由于含有错误而无法使用，甚至带来巨大损失，软件给人的感觉是越来越不可靠，以致几乎没有不出错的软件。这一切，极大地震动了计算机界，史称“软件危机”。人们认识到：大型程序的编制不同于写小程序，它应该是一项新的技术，应该像处理工程一样处理软件研制的全过程。程序的设计应易于保证正确性，并且便于检验。1969年，提出了结构化程序设计方法，1970年，第一个结构化程序设计语言——Pascal语言出现，标志着结构化程序设计时期的开始。从此计算机程序可以让程序员以接近问题本质的方式去思考和描述问题。

（三）面向对象的软件编程语言——以人类思考问题方式解决问题

计算机高级语言产生后，虽然可以支持类似于人类的思考方式去编写程序，但是这阶段主要采用面向过程的方法，把问题按照需要解决的步骤，进行一步一步的处理，这种方法虽然解决了问题，但是过于抽象，在一个模块被执行完成前，无法动态地改变程序的执行方向。这和人们日常处理事物的思维和习惯并不一致，人们日常的习惯是形象化、直观化看待事物，发生一件事就处理一件事，对于计算机软件来说，人

们使用时是面向具体的应用，并不过多关心过程。具体的应用就是对象（object），面向具体应用的开发过程就是面向对象开发。面向对象的开发思想，让计算机能够尽可能以人的自然思维方式去处理问题，强调了以自然世界的对象及相互关系为中心来思考、认识和解决问题。

面向对象的思想产生以后，计算机的思维开始以更加接近人类的思维方式解决问题。面向对象的编程的高级语言主要有：C++、Java等。

高级语言的下一个发展目标是面向应用，也就是说，只需要告诉程序你要干什么，程序就能自动生成算法，自动进行处理，这就是非过程化的程序语言。计算机语言和程序设计开始向人工智能方向挺进。

第三节 人工智能

人工智能（Artificial Intelligence，简称AI）是计算机学科的一个分支，它实现的主要物质基础以及能够实现人工智能技术平台的机器就是计算机，人工智能的发展是伴随着计算机科学技术的发展而成长起来的。人工智能的发展方向有很多，主要研究领域强调模仿人类思维的学习、推理、思考、规划等活动。人工智能的发展方向简单地说，可以分为符号智能和计算智能，符号智能指以符号形式的知识和信息为基础，通过逻辑推理运营知识进行求解。主要包括知识获取、知识表示、知识组织与管理、知识运用等技术；计算智能是以数值为基础，通过数值计算和运用算法对问题求解，主要包括神经计算、遗传计算、进化计算、免疫计算、粒群计算等。符号智能技术与图书馆的知识管理密切相关，图书馆需要对以文字符号编码的人类智慧进行管理和利用。在图书馆领域普遍对人工智能的理解，更加倾向于美国斯坦福大学人工智能研究中心尼尔逊教授对人工智能的定义：“人工智能是关于知识的学科——怎