

办公信息系统中的计算机科学技术问题

王能斌 黄大海 冷嘉玲

(东南大学)

摘要

计算机应用是计算机科学技术发展的动力和目的。在当前诸多计算机应用领域中,最有代表性的是基于知识的系统(KBS)、计算机集成制造系统(CIMS)、办公信息系统(OIS)和实时控制、指挥、通信系统(C³IS)。它们向计算机科学技术提出一系列富有挑战性的课题,成为当前计算机科学技术研究的热点。本文仅就办公信息系统所提出的计算机科学技术问题,做一概述。

办公信息系统不仅有其技术上的特点,还有其社会的属性;它不仅涉及计算机科学技术,而且还与系统工程、知识工程,人机工程、管理科学、经济学、社会学、人类学、心理学等多种学科有关^[1]。本文不拟对此有所论及。办公信息系统与其他计算机应用系统一样,是建立在计算机硬、软件和通信、办公设备发展的基础上。对于这些物质基础,办公信息系统虽有其自身的要求,但与其他系统基本上是共同的。事实上,办公信息系统一般是用这些设备的商品化产品构成的,故本文也不拟对此进行讨论。办公信息系统是一种综合的计算机应用系统,要讨论其发展中的所有问题是困难的,下面择其

主要问题,谈谈作者的看法。

1. 联邦式的系统结构 (Federated Architecture)

— 办公室一般分为多个部门。这些部门往往在地理上是分散的,在组织上既相对独立,又互相联系、彼此合作地工作,而且还可能有派出人员,分布在世界各地。在这样的系统中,数据的来源是分散的,既有供部门自用的私有数据,也有允许其他部门使用的共享数据,甚至允许不同部门拥有重复且不一致的数据。目前已开发的分布式数据库绝大部分在物理上是分布的,在逻辑上是集中的,即把地理上分散的数据综合成一个全局数据模式,由DBA集中管理。这种形式的数

• 通过重用软部件,开发和使用应用程序生成器以及第四代语言,避免软件镀金现象等来减少代码的编写。

作为最后的结论,有一点值得特别强调。在回顾和追踪软件生产率提高的过程中,我们必须注意不要把方法与目的相混淆。提高软件生产率本身并不是目的,它是帮助人们更好地扩大处理信息和做出决策的能力的一种方法。通常帮助别人做些事情时

将使我们牵连到各种活动中(如花两个星期帮某人找到一个问题的非软件的有效解答),而这些活动并不增加我们在生产率计分牌上的分数。我们必须时时记住软件生产率这块计分牌只是我们用计算机更好地为人民服务方面的进步的许多尺度中的一个。

参考文献略

[杨德元、杨军译自《Computer》1987, No.9, 王启鸣校]

据管理显然很难适应办公信息系统的要求。姑且不论全局数据模式设计和管理的复杂性,即使构成了一个全局数据模式,也难以满足用户各式各样的需求以及适应灵活多变的办公环境。事实上,在大型办公系统中,几乎没有一个人能了解全系统的情况和需求。在办公室环境中,数据的不一致是经常发生的,而且在某种情况下是允许的。这绝不是DBA所能管理得了的。R*首先在分布式数据库系统中提出并实现了结点自治,这是一大进步。近几年又提出了联邦式系统结构,在这个方向又前进了一步^[2]。联邦式数据库由一组松散联合的独立的数据库组成,它们可以通过协商互相交换、共享数据。每个独立的数据库有三种概念模式:

- **私有数据模式**(private schema) 描述为本结点所用的局部数据。

- **输出数据模式**(export schema) 描述本结点愿与联邦中其他成员共享的数据,并规定访问权限(例如允许读或写)。

- **输入数据模式**(import schema) 表示本结点拟用的其他结点的数据,可以从有关结点的输出数据模式推导而来。

通过输入、输出数据模式,各结点可以根据需要共享其他结点的数据,而不必求助于全局模式。利用其他结点的输出数据模式,建立本结点的输入数据模式,是个双边或多边的协商过程。为此,在联邦式系统结构中,设有协商机制,提供一种协商语言及其相应的解释器,但系统中无全局性的控制机构,各结点可以自治而合作地工作,充分反映了办公室的特点。

2. 多媒介信息的管理和处理

在现代化的办公室里,信息不但以文字形式表示,也常常以声音、图形以及图象等形式表示。这称为多媒介信息。过去由于受到存贮量及计算网带宽的限制,未获广泛应用。近年来,由于光盘和宽带局域网的商品化,多媒介信息的应用逐步得到发展,成为现代化的办公信息系统的组成部分。在同一

文件(document)中,往往这几种信息表示形式并存。例如在一个文件中,既有各种属性,又有文字、图形和图象,还附有声音说明。它们所占的存贮量和处理要求差别很大,物理表示形式更是千差万别,用传统的格式化的数据模型很难表示这样的信息。采用面向对象(object-oriented)的数据模型可以比较满意地解决这个问题。面向对象的数据模型是抽象数据的进一步发展。它的基本思想仍然是封闭和信息隐蔽(encapsulation and information hiding),即把形形色色的信息统统抽象成对象,用户看到的仅仅是它的外部属性、与其他对象的关系以及对它的操作,而信息的表示形式、数据结构、操作过程等用户不须关心的内容,一概对用户隐蔽,由系统实现和维护。目前已有一些实验性的面向对象的数据管理系统出现,例如ORION^[3]、POSTGRES^[4]等,商品化产品的出现也为期不远。面向对象的数据管理系统将是多媒介信息管理及处理的主要手段,在现代化的办公信息系统中将获得广泛应用。

多媒介信息和多功能多窗口的用户接口相结合,将使办公环境大为改观,这是办公信息系统发展的方向之一。

3. 非结构化办公例程和协商过程的处理

早期的办公信息系统主要是代替人做一些办公室的例行事务,例如文件的编辑、存贮、传递、分发、检索及报表、统计、预测、模拟等任务。这些工作都是结构化的问题,比较容易实现。随着办公室现代化程度的提高,办公信息系统也面临着许多非结构化的问题,例如某些非结构化的决策问题,裁决问题等。这些问题的解决需要以知识为基础。因此,各式各样的基于知识的系统已在办公信息系统中得到应用,例如COKE^[5]就是一个办公知识工程系统,用它可以实现智能办公助理(knowledge-based office assistant)。这种办公助理具有办公室的人

事、组织、管理等知识,可以为领导完成准备月报、分配经费之类的任务。

办公室采用基于知识的系统后,带来了新的问题,就是各部门的知识不一定一致。要取得知识的全局一致,既不可能,也不必要。事实上,在现实的办公室中,由于观点不同,历史背景不同,利益不同,制度不同,处境不同,常常发生这样或那样的分歧。即使原来没有这些分歧和差别,在办公室环境中,也会发生一些偶然的或例外的事件,又可能引起新的分歧。聪明的领导不是企求没有分歧,也不是主观地去裁决,而是了解各方面的难处,通过协商、辩论,求得一个合理的、可行的解决方法,使得各个部门能够贯彻实施。这个过程叫做协商过程(due process)^[6]。协商过程不是靠裁决,而是在诸矛盾要求中,寻找可行的合理的解。在办公室工作中,不一定非要等所有信息全部到齐了才做决定。这样做的决定可能和迟到的信息和偶发的事件发生矛盾。解决这些矛盾也可借助于协商过程。

4. 办公信息系统的一些通信要求

这里不准备讨论计算机网络的发展,这是另一门学科的内容。办公信息系统对计算机网络的发展虽有一定的促进,但仅仅是诸因素之一。不可能从一个应用系统的要求推测计算机网络的发展。目前虽有一些面向办公信息系统的网络标准(例如TOP)和产品,但毕竟是计算机网络技术发展的产物。计算机网络是办公信息系统的重要物质基础,办公信息系统的发展有赖于计算机网络技术(例如ISDN)的发展,这是毫无疑问的。

本文着重从应用的角度谈谈办公信息系统对通信的要求,也就是主要讨论有关网络应用层的问题,这是办公信息系统的组成部份。下面列举办办公信息系统对通信的一些要求^[7]。

• **速度** 办公信息系统对有些信息的获得有一定的时间要求,例如外派销售人员要查询仓库中是否有货可以供应,如果未能及时得到这个信息,则

将丢失一笔生意。再如秘书用电子日历(electronic calendar)安排会议,必须在规定时间内完成,而这种安排工作往往需要大量的通信。

• **信息的完备性** 办公室在作决定或处理问题时,总是希望获得完备的有关信息,否则会造成失误。例如在作销售预测时,希望能够获得所有地区和销售点的预测报告。在安排会议时,希望不要漏掉任何一个有关人员等等。

• **把信息送到应访问的对象** 在办公系统中,有时只有收信人的条件,而无明确的收信人的名单。系统须将信件送至应收人手中。有时还须通过多次向其他单位或个人间接访问,才能确定收信人。在此情况下,漏送固然不好,多送了也不好,既增加了通信的开销,也增加了收信人的麻烦,也不符合保密的原则。

• **对偶发事件的反应能力** 在办公系统中,常常会有临时发生的偶然事件。例如一个值班人员突然生病了,一个应该到会的人临时因故不能出席会议等等。办公信息系统应能将这种信息主动送至各有关部门,作相应的处理。

• **传递环境信息** 在办公系统中,有时不但要传递文字信息,还希望传递现场的音、像、图、景以及笔迹等,以增强效果。

为达到上述要求,须发展智能邮件系统和多媒体部件系统。

5. 友好的、多样化的用户接口

用户接口历来是办公信息系统研究的重点。这直接影响到办公信息系统的推广与应用。多窗口、图标(icon)、菜单等已经普遍应用。进一步的发展将是采用手写体输入、语音识别、自然语言理解等技术,使系统能与用户以人们所熟悉的自然方式进行对话。

在办公信息系统中采用自然语言固然是发展方向,但在短期内很难实用化。发展友好的办公语言仍然是重要的。这些语言应尽可能与自然语言接近,适合办公人员的习惯,易学易用。

汉字处理是我国发展办公信息系统的关键问题之一。现在虽然发展了很多的汉字输入方法,其中有些方法的输入速度也不亚于西文。但使用人员毕竟还要受过一定的训练,这仍然是推广办公信息系统的障碍。发

展汉字手写体输入和汉语语音输入是解决这个问题的重要途径。汉语自然语言理解有其特殊问题,需要专门研究。语言的理解除了需要语言知识外,还需要背景知识。要解决通用化的汉语自然语言理解是困难的,比较现实的途径是发展用于某一领域的汉语自然语言理解。也可以发展一种通用的汉语自然语言理解框架,加入不同的背景知识后,可以生成用于不同领域的汉语自然语言理解系统。

6. 用户介入开发与维护

国内、外的经验都证明,按照软件生存期分阶段开发办公信息系统是不可行的。其理由有三:(1)最了解办公过程的是办公人员。他们在没有使用办公信息系统之前,很难提出完整的要求。而在系统交付使用时,他们才发现系统的不足,甚至拒绝使用。(2)办公环境经常变化(组织变化、人员变化、政策制度变化、功能变化等等),很难想象,根据几年前的需求分析所设计的系统,能满足几年后办公的需要。办公人员常常因办公信息系统的功能不符合实际情况而拒绝使用。(3)办公室中常常会出现一些临时的、紧急的任务,而这些任务在办公信息系统设计时并未考虑,因而无法完成。用户往往因此而对办公信息系统感到失望。解决这个问题的出路在于用户直接参加系统的开发和维护。正象汽车那样,不是由专职司机驾驶,而是由乘车人自己驾驶。速成原型法就是为了迎合这样的需要而发展起来的一种开发方法。设计者可以根据用户的初步需求,采用各种开发工具、非过程语言以及现成的软件及模型等手段,快速地构成系统的原型,由用户来操作使用,反复地进行修改,直至用户满意为止。然后再在原型的基础上,发展成实用系统。速成原型法主要解决了需求分析的正确性问题以及用户的合作问题。为了使用户能够直接维护办公信息系统,还必须研究各种面向用户的程序设计语言和开发工具以及简明的指导文件。总之,办公信息系

统应该是这样的计算机应用系统,在它的基本骨架上,办公人员可以根据需要经常地进行修改和补充。

7. 不同的新老硬软设备的集成

办公信息系统常常是由小到大逐步发展的。由于经费的限制,办公设备不可能像一般计算机那样,几年更新一次。因此,在同一办公系统中,往往不同厂家的设备以及新老设备并存,要把这些设备集成为一个系统,是一个很复杂的技术问题。当然,最根本的出路在于标准化。但是这需要一个漫长的发展过程。作为一个过渡办法,可以采用各种转换技术。在硬件、软件和通信中,都有各自的转换技术,本文不拟论述。就办公信息系统的设计而言,可以采用两方面的措施:一是尽可能减少转换次数(例如都转换成统一的系统标准,而不是双边转换);二是在设计时,尽可能考虑到转换的方便,甚至事先就设计好各种转换接口。

以上所讨论的问题不过是作者认为比较重要的一些问题,远未概括办公信息系统发展中所涉及的所有计算机科学技术问题。有些问题,例如保密问题,在办公信息系统中也很重要,但由于在其他领域中,对此问题研究得更多,就没有列入。文中有些观点是作者的一些看法,难免有片面和疏漏之处,欢迎读者批评指正。

参考文献

- [1] C. Hewitt, Offices Are Open Systems, ACM TOOLS 4:3(1986)
- [2] D. Heimbigner, D. Mcleod, A Federated Architecture for Information Management, ACM TOOLS 3:3(1985)
- [3] D. Woelk, W. Kim, Multimedia Information Management in an Object-Oriented Database System, Proc. of 13th VLDB, 1987
- [4] M. R. Stonebraker et al., The POSTGRES Data Model, Proc. of 13th VLDB, 1987
- [5] A. R. Kaye, G. M. Karam, Cooperating Knowledge-based Assistants for the

新一代数据管理系统

周龙骧 (中国科学院数学研究所)

摘要

集中式数据库管理系统在十年前已经推出了若干成功的商品。分布式数据库系统经过十年的研究与开发,其技术已经成熟,正处于已经推出和正在推出一些商品的时期。数据库管理系统已经进入了酝酿、研究和开发新一代系统的时刻。

一、引言

衡量一门学科是否已经成熟,其基本问题是否已大部分得到解决,可以看它是否已出了一本比较完整的书。衡量一种计算机软件技术是否已经成熟,可以看它是否已有一种提供市场的产品推出来。集中式的关系数据库系统十年前已经推出了若干种成功的商品,因此可以说关系数据库技术十年前已经成熟了。分布式数据库技术从七十年代下半期开始,投入了巨额的资金和研究努力,经过十年的研究和开发,已有若干原型系统研制成功,并已推出和正在推出一些商品。因此可以说分布式数据库系统技术也已臻于成熟。从八十年代初开始(甚至可以上溯到七十年代末)人们开始探讨数据库系统新的研究方向。新的DBS样式(Shape)(例如数据库机,数据库和人工智能的结合——演绎数据库,知识库,专家系统等,数据库和计算机网络的结合即分布式数据库);数据库的新的应用领域(例如管理信息系统,办公自动化系统,决策支持系统, CAD/CAM/CASE/CIM);

和数据库的新应用密切相关的是数据库作为大系统的一个组成部分的重要作用;新的数据模型;新的数据表示和结构(如图形,图象,声音,自然语言,正文、文档等);新的用户接口以及数据库理论的深化。这些新方向的研究取得了丰硕的成果,它们深化和提高了原有的理论和技术,使原有的系统更为精炼,高效,应用更为广泛,同时开拓了未来,迎来了新一代数据管理系统的研究和开发时代。

二、历史回顾

虽然早期的高级语言大多数是以程序为中心,如ALGOL, FORTRAN等,但数据处理却扮演着越来越重要的角色。回顾起来数据管理已经走过了四个阶段,有了四代数据管理系统。

1. 第一代(1950年代) 这一代DMS基于顺序文件,其硬件基础是磁带,著名的语言是COBOL。数据按记录组织,顺序记在顺序文件(即磁带)上。这一代DMS是面向批处理的。

2. 第二代(1960年代) 由于出现了分

Office, ACM TOOLS 5:4(1987)

[6] E. M. Gerson, S. L. Star, Analyzing Due Process in the Workplace, ACM TOOLS 4:3(1986)

[7] S. F. Ehrlich, Strategies for Encouraging Successful Adoption of Office Communication Systems, ACM TOOLS 5:4 (1987)