

人工智能上

习情报检索

贾同兴 编著

人工智能与情报检索

贾同兴 编著

北京图书馆出版社

图书在版编目(CIP)数据

人工智能与情报检索/贾同兴编著:—北京:北京图书馆出版社,1997

ISBN 7-5013-1341-5

I.人… II.贾… III.人工智能-关系-情报检索-研究

IV.G354

中国版本图书馆 CIP 数据核字(97)第 10746 号

书名 人工智能与情报检索

著者 贾同兴 主编

出版 北京图书馆出版社(原书目文献出版社)

发行 (100034 北京西城区文津路 7 号)

经销 新华书店

印刷 北京航空航天大学印刷厂

开本 850×1168 毫米 1/32

印张 8.5

字数 200(千字)

版次 1997 年 7 月第 1 版 1997 年 7 月第 1 次印刷

印数 1-3000 册

书号 ISBN 7-5013-1341-5/G·367

定价 12.80 元

前 言

认知科学、信息科学和管理科学的发展，更新了人们的科学观和方法论。计算机、通信、知识库、多媒体及人工智能技术的进步，日益改变着人们的研究、设计、生产、决策、学习、生活和思维方式。在这种形势下，当我们用更高的理性思考传统的情报检索时，就不难发现在理论和实践方面，传统情报检索都面临着“重新构筑、重新设计、重新培训、重新起步”^{〔78〕}的挑战。

“人工智能与情报检索”正是在这种背景下产生。在第一章“绪论”中，从认知科学的高度评述了传统情报检索的功过，揭示了人工智能与情报检索在实质上的等同关系，提出了情报系统框架。

第二章“人工智能”中，概括地探讨了人工智能 40 年来的成就、研究现状及存在的问题，指出了继续研究的路线，同时深入浅出地介绍了人工智能的基本假设、基本知识和技术，为第三章“情报检索”提供反思的背景。

第三章“情报检索”中，深入地探讨了情报检索的超文本技术、情报检索的自然语言处理、知识库的组织技术、并行情报检索、情报检索的不确定性及情报检索的理论基础。

第四章是国外 3 个处于实验原型阶段的智能情报检索系统，它们被引率很高，对我们设计智能情报检索系统很有参考价值。

第五章探讨了情报检索的基础理论学科，介绍了它们的基本理论、研究对象、研究现状及展望，同时介绍了基本知识，提出了情报检索的知识结构图。

本书是作者 10 年研究人工智能与情报检索的成果。作为一部学术著作，它应完成两大学术任务^[30]：1. “使后继者获得较为正确的知识和方法，迅速把他引到探索的前沿；2. 它应该指出作者还来不及走但应该走的路，即提出这门学科未来的新的生长点”。本书谨守此宗旨。

我国智能情报检索研究还处在起步阶段，作者希望该书能给读者介绍些有用的知识，解决点问题，更希望引起国内同行的注意和对人工智能的研究兴趣；同时，也渴望人工智能界的专家能投身情报检索的研究，以促进情报检索迅速健康的发展，若能如愿，将是对作者 10 年辛勤劳动的最好报偿。

作者 1996 年 4 月

目 录

前 言	(1)
第一章 绪论	(1)
1·1 现代信息系统的五元组成	(1)
1·2 现代信息技术	(4)
1·3 联机情报检索的贡献、问题及对策	(15)
1·4 新编情报检索技术大事记及述评	(23)
1·5 情报检索定义新说	(36)
1·6 人工智能与情报检索	(40)
第二章 人工智能	(49)
2·1 人工智能简史	(49)
2·2 人工智能研究综述	(53)
2·3 人工智能与知识工程	(85)
2·4 从哲学看人工智能和从人工智能看哲学	(87)
第三章 情报检索	(96)
3·1 国外近 10 年的智能情报检索研究述评	(96)
3·2 情报检索与自然语言理解	(110)
3·3 智能情报检索中的超文本技术	(116)
3·4 情报检索中的信息组织	(125)
3·5 信息管理中的不确定性	(129)
3·6 并行处理与情报检索	(139)
3·7 两级概念结构——情报检索技术的重大革新	(149)
3·8 认知科学与情报检索	(158)
第四章 智能情报检索的三个实例 (附录)	(165)

4 • 1	智能中间人情报检索(I ³ R): 一个文献检索系统设计的新方法	(165)
4 • 2	概念情报系统的知识组织与存取 LISA F • RAU	(191)
4 • 3	情报检索专家系统的原理、过程和规则	(213)
第五章	情报检索的理论基础	(231)
5 • 1	信息检索的知识结构	(231)
5 • 2	认知科学	(232)
5 • 3	信息科学	(241)
5 • 4	管理科学	(247)
5 • 5	脑科学	(250)

第一章 绪 论

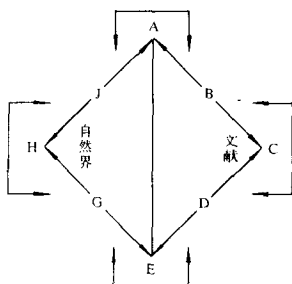
1·1 现代信息系统的五元组成

人类社会，正由工业社会迈向信息社会。信息社会的工作、生活、思维模式正展现在人类面前。西方发达国家的社会现实，已经为发展中国家显出了这种模式框架。在我国的一些大城市也初露该框架的端倪，我们这里初步勾画一个粗框架，以便为我们的讨论提供论域。这种模式由 5 个部分组成，即信息、计算机、网络、信息库和人。

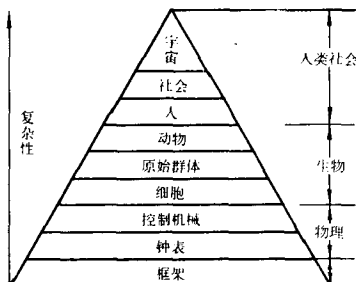
1. 信息

这是计算机加工的、通信网中传递的、信息库中组织的、人机交流控制和在脑内知识建构的对象。信息可以是人脑中的信息，也可以是自然界或人造系统的信息，请参照下面信息交流和分类图（见下页）。

从以下两图我们可以看到信息交流领域的划分及信息的层次，正是信息构成高级社会的资源，是物质、能源以外的第三种资源，这些信息或单独或在库中或在自然界中通过使用流动，它就象货币一样，流动起来就可以增值，产生财富，推动社会进步。



信息交流模型*



信息的系统分类**

2. 计算机

它是一种现代社会不可或缺的信息加工工具。计算机网可以由巨型、大型、中、小、微型机组成（开放系统互联网中的分布网），微机可以是家庭或办公室中的节点机，中小型机可以是具有各种数据库、知识库，能作复杂处理的节点机，大、中、巨型机可以是并行和共享的。进一步发展可以是智能机。智能计算机系统有丰富的系统软件和各种应用软件、智能软件、学习软件等供资源共享。

3. 开放系统互联网

INTERNET 网就是一种雏型。这是一种由国际标准化组织开发的，一种能使一组开放的系统合作的网络化、标准化结构模

* 见 B.C. 维克利等《情报科学的理论与实践》，科技文献出版社，37 页

** 见薛华成《管理信息系统》，清华大学出版社，44 页

型，这些系统通过联合使用适当的 ISO 标准而实现信息交换。这种网间通讯必须遵守 ISO 制定的协议通信。这种网中网，还可以包容各种分布性网、计算机集成制造系统网和大众媒体传播系统网等。在这种网络中，任何人，任何时间都能与任何地方的另一个人进行通讯，所提供的业务仅受接入网或终点网及用户终端能力的限制，而最终将向用户提供任何形式的信息服务。这时才真叫秀才不出门，便知天下事。

4. 信息库

计算机和网络所以能发挥作用，主要靠信息库、知识库，与库内信息的优质和严密的组织结构、信息库的组织情况及数据存取效率关系极大，尤其是在多媒体情况下，各种关系错综复杂，交叉引用。互联网是个非常复杂的巨系统，在与系统软件的配合下，它能及时更新，能自动建库，保持库的活力，给问题求解以巨大支持。

5. 人文因素

这是任何系统的主导，是系统智力的源泉。它与系统形成人机联作，形成各种各样的互补。人机共建，共存共荣，遵守拟人、辅人和人机共生规律⁽²⁾，合作解决难题。比如，目前智能系统知识少，遇到尴尬局面就需请人类专家帮助，而由机器做比人效率高的工作，如高速运行的机器的控制。高温、低温、有毒、有危险的环境，人也都可以避开。总之，任何智能系统都离不开社会性、人文性，系统是人设计的，用户是人。处理的对象与人类息息相关，人要在此系统中发挥能动作用，还需注意人机工效学的理论，以提高人机效率，人还得安全避害、舒服，系统要降低消耗，以上五元素是现代信息系统的主要组成部分。

1·2 现代信息技术

现代信息技术的内容极为丰富,我们主要讨论计算机技术、通信技术和数据库技术。

1. 计算机科学技术

我们在讨论人工智能和情报检索的关系时,无论从哪一方面,似乎都应该谈谈计算机技术,因为它不仅是双方使用的工具,而且是维系当代社会的“金三角”中的一角(即计算机、数据库和人类智慧)。数据库或知识库是人类几千年智慧的结晶,计算机系统(包括人工智能系统)是人类智能的延伸,将来可能作为人的伙伴与人“平起平坐”(当然它还是受人支配的人造物),人类智能是三角中的控制因素,无疑最重要。当代世界三元素之间由人造的通信网联起来,成为信息的有机的整体系统。

1) 计算机的历史

计算机的历史简单地可分两个阶段:冯·诺曼机和非冯·诺曼机(智能机)。在冯·诺曼机阶段(1946—1982),从1954年起始的用计算机处理文字,是计算机应用史上的里程碑事件。60年代后的计算机与通信网的结合具有非常重大的意义。发展到今天的开放系统互联网,加上计算机会议、多媒体技术,给历史上的文化积淀与现代人头脑中的知识交流创造了无与伦比的科技环境,改变了人类研究、设计、决策、工作和思维的方式,为人类高效工作立下了不世之功。1981年,日本宣布的第五代机研制计划虽未成功,但对智能机的研制工作起了极大的推动作用。智能机的研究,受勿需说明的智能转移规律(辅人率、拟人率与人机

共生率)影响,早晚必将获得成功。

2) 计算机、人工智能和情报检索

计算机是一种信息处理工具,这是当初研究创造计算机的人所没有想到的。人工智能、情报检索与计算机结合,都是借助计算机来揭示人类信息加工的奥秘,并进而把人不胜其烦的信息处理转移到智能机系统上。这在计算机与人工智能、情报检索关系上,真可以说“一拍即合”,就目前人脑研究的成就看,人脑主要功能就是信息加工。计算机目前的水平也只能作信息加工工作。所以说计算机找到了用武之地。人工智能和情报检索找到了最合适的工具。

3) 未来的计算机——智能机

未来的计算机一定是智能机,具有解题和推理、知识库管理和智能人机接口3大功能。目前智能机的研制主要从神经计算机和光计算机入手。

①神经计算机。神经计算机的研究主要从两方面入手,一个是从低等的动物神经细胞(如线虫)试制人造脑机,对人造脑的研究还处于基础研究阶段。另一个是在硬件和软件上下功夫,通过仿效人脑使之具有人脑功能。目前,研究已取得了一些成果,如日立公司开发的在一个硅片上集成144个神经元,并由多个硅片联成的神经计算机;美国人研制的识别人的脸谱的计算机,可识别500人的脸谱。

②光计算机^[4]。是把光学器件和处理元件制作在同一块硅片上的光学电子混合系统。由于光学处理元件通过光束联系,可以克服硅片上布线的物理极限问题。如美国加州理工学院的实验装置,包括一个由1万多个神经元构成的二维阵列阈器件,一个由透镜、反射镜系统、针孔阵列和两张平面全息图组成。阈器件上

阈元件接受来自背面射来的光束，当光束大于某一阈值时，就使其正面反射能力发生改变，它起神经元的作用。透镜、反射镜和针孔阵列规定每一阈元件所得光的多少，实际上建立了这些元件间的关系：两张全息图含有同一组图象，但其中一张图象的边缘作了增强处理，整个系统布置成光学回路形式，使之可以产生连续反馈。这就是光学计算机，我们利用它识别图象的过程如下：一个要识别的图象被阈器件的正面反射后，投射到模式识别系统中。透镜、反射镜和针孔阵列使反射的输入图象，能与存贮在两张全息图中的所有图象以下述方式相互作用，使从第二张全息图上出来的最明亮的图象，就是输入图象与存贮的全息图之间的最佳配合图象。然后，从第二张全息图上出来的光射到阈器件的背面，引起各个阈元件的反射能力发生变化，从而使最佳配合的图象，基本上就是从阈器件的正面反射回去而绕回路进行第二次循环的图象，通过回路中连续循环，使最佳配合的图象不断增强，直至找到存贮的图象。

③日本的第五代机计划。要完全生产神经和光计算机，目前还是一种理想，但据目前取得的研究成果和计算机生产的实践来看，造出更聪明的计算机应不是太远的事。

70年代中期以来，发达国家的决策人已普遍认识到：信息革命正改变着国家之间财富与实力的对比，而信息革命，首先是计算机技术的革命。所以出现了美国的战略防御计划、西欧的尤里卡计划及日本的第五代机计划。

日本人清醒的看到，只有头脑是自己拥有的财富，所以决心改变“模仿者”的二流国家形象。想以五代机的研制为契机，进而掌握 21 世纪的科学技术制导权。

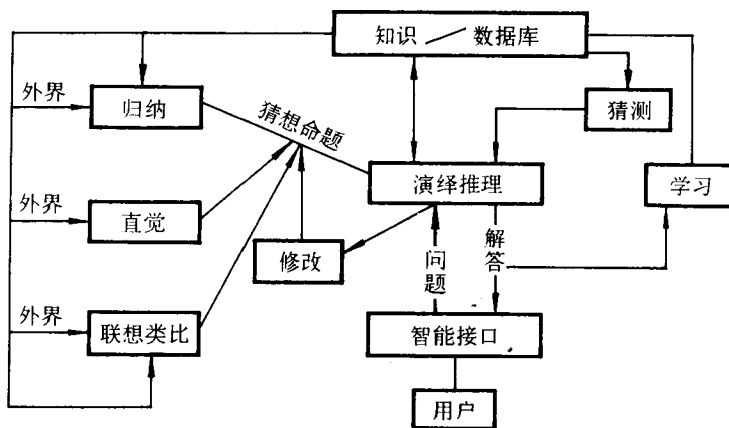
日本人的五代机的主要研究课题如下：

a. 基本应用系统：机器翻译，回答系统，应用语言理解，应用图形、图像理解与问题求解；

- b. 软件系统：知识库管理，问题求解与推理，接口系统；
- c. 新的高级结构：逻辑程序机，数据流机等；
- d. 分布式功能结构：分布式功能结构及高级人机通信机器，大规模集成技术；
- e. 系统技术：智能编程系统，知识库设计系统及数据库与分布式数据库技术；
- f. 开发支持。

从以上介绍看，除了支持和创造技术外，可以认为，五代机就是知识处理系统。与情报检索系统同质、同构。

有文献^[3]指出了智能机的几大功能：能存贮大量信息和知识；会思维和推理；能进行高速并行处理；能用自然语言、文字或图象等和人进行信息和知识的交流；能试探求解问题和有一定创造力；具有学习功能和能不断修正完善自身；能解决多种问题和只需告诉它作什么，就可以自动工作。下面的图 2 是一种智能机示意图^[35]：



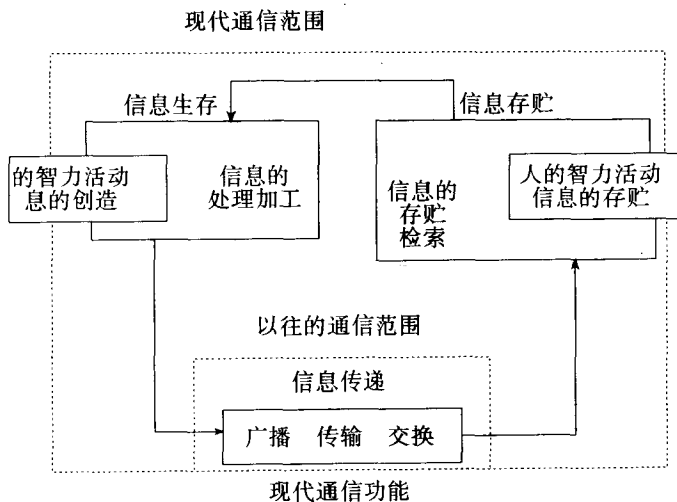
智能计算机示意图

2. 现代通信技术

1) 概论

在信息已成为最重要的战略资源的今天，如果说计算机系统就是现代社会的大脑，那么通信系统就是现代社会的神经系统。计算机与通信结合，才能成为“完人”，才能完成现代信息的战略资源使命。

现代通信可以认为是由各种媒体表达组成的，这些表达是对通信双方共有的一个知识结构产生激发，从而完成信息传递。现代通信的功能如下图：



现代通信的介质有：声、光、电、磁。通信技术有：卫星通信、光纤通信。通信的系统可有公用系统、事物系统和家庭系统。这些系统又通过网络联在一起。网有局域网、远程网，这些网又通过协议，形成开放系统互联，如：INTERNET 网。现代社会的

许多活动都在家庭中进行。

2) 家庭信息系统

下面，我们就以家庭系统为例，描述一下通信情况：

由于家庭办公及生活、娱乐信息的处理，家庭系统是未来系统网中的一个极为重要的终端系统。但由于人类社会正进入信息时代，还不能进行非常具体的描述，但大的框架一般可能如下：

①家庭系统的信息^[9]。家庭系统的信息主要有四类：

- 娱乐信息。家庭是人们休息的地方，各种娱乐的信息将是最主要的、大量的信息。

- 生活信息。如新闻、天气预报、经济及行情、灾害等。

- 生活行动信息化方面的信息。如购物、利用金融机构办理各种手续，预约定购车、船、机、院的票。

- 在家办公信息。由于创造性劳动和家庭主妇工作要求的需要，办公信息也将包括在家庭系统中。

②家庭系统的组成。它由 3 个子系统组成：家庭控制、家庭生活和家庭工作。

- 家庭控制子系统。居室内各部位的监视器、冷暖气、门的开闭、电源切换及报警系统。

- 家庭生活子系统。收音机、电视机、新功能电话、可视数据（如用户智能电报）、CATV（有线电视）业务咨询，家庭购物存款。

- 家庭工作子系统。信息处理、文字处理，访问企业信息处理中心的通信功能、记录通信的功能。未来的家庭系统中，将有以家庭计算机为中心的各种信息业务终端，并与家庭网络（点线）相联，然后接入各种局域网。

③家庭系统及业务动向。家庭系统的设备主要是电视机、电话、个人计算机、收录机等。这些设备组合在一起就变成各种媒

体终端，其中计算机将管理出入家庭和家庭内的信息。它们有加工处理，存贮检索及文字处理功能，计算机将分别适合家庭不同层次成员的选择。如家庭办公用的事务处理机，以处理家务的主妇用的计算机，学习用的儿童机等。

④家庭系统用的接入系统。接入家庭系统的通信系统包括：无线系统、有线系统和标准件系统。

a. 无线系统。目前包括无线电广播、电视广播及直接广播卫星系统（DBS），移动通信系统，它们是家庭通信的延伸。

b. 有线系统。主要是有线电视（CATV）广播（目前主要使用电缆，将来会用光缆）和模拟电话系统。正在研制的综合业务数字网（ISDN）用户接入系统，以 2B+D 为前提，使用标准传输速度的接口，所谓 2B，指的是两条双向的 64kbpsB 信道和一条 16kbpsD 信道，这种接入系统的实现，可以实现图象、声音、文字的综合传输，输送将用光缆，并且将与广播业务联结起来。

c. 标准件系统。实际是指磁带、磁盘等标准件信息供给系统。这是一种离线系统，在娱乐教育领域占很大比例，录音和录象由于实现了数字化，其音质和画面质量将有很大提高。录象、摄象合为一体的高档音响和图象设备，随着卫星广播的实际应用，将会逐步普及。

以上我们通过家庭系统这种通信形式介绍了现代通信技术的各个方面。总之，通信与计算机技术互补形成信息技术的核心技术，是现代社会的五元组极为重要的两元。

3. 数据库知识库技术

数据库技术是有关计算机数据管理的一门技术，利用数据对科研、设计、生产进行管理以及管理科研、设计、生产中积累的数据、信息和知识，是发展科学技术，促进社会进步的极为重要的因素，尤其是能把历代积累的人类文明（静态知识）与活人头