



形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

形式语言与自动机理论

姚 刚

中国科学院信息工程研究所



目录

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

- ① 课程简介
- ② 希尔伯特的第十问题
- ③ 自动机研究历史
- ④ 形式语言研究历史
- ⑤ 课程中的概念



课程简介

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- 课程属性： 专业普及课
- 学时学分： 40学时/ 2学分
- 考核方式： 课堂开卷
- 主 讲 人： 姚 刚
- 上课时间： 星期五， 第5、6、7节
- 上课地点： 雁栖湖园区教1-123



参考教材

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- ① Peter Linz. An Introduction to Formal Languages and Automata.
- ② John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman. Introduction to Automata Theory, Languages and Computation.
- ③ 陈文字, 田玲, 程伟, 刘贵松. 有限自动机理论.



课程目标

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- 熟悉计算机科学的基础和原则，规范推理能力；
- 了解可行程序或算法的分类及特征；
- 了解计算模型；
- 熟悉可解问题和不可解问题。



课程内容

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- 自动机、语言、文法的概念；
- 有限自动机及其类型；状态化简；
- 正则语言及其性质；
- 上下文无关语言及文法；
- 下推自动机的概念；
- 图灵机的概念及其模型；
- 形式语言与自动机的层次；
- 计算模型、计算复杂性、量子自动机等。



问题起源

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- 数学问题是数学中最具魅力的一部分。
- 第二届数学家大会，1900年，巴黎
- 希尔伯特(Hilbert)的演讲—数学问题，列举了23个他认为最具重要意义的数学问题。



丢番图(Diophantus)方程

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究历史

课程中的概念

人们对有如下特征的方程感兴趣：只包含未知数的整数次幂，系数也都是整数的方程。例如： $2x - 4y = 5$ ， $x^2 + y^2 = z^2$ 等。

对于丢番图方程，人们最感兴趣的问题是这类方程是否有整数解？对于一般的丢番图方程，判断它是否有整数解是一件极其困难的事情。例如著名的费尔马(Fermat)猜想。



希尔伯特的第十问题

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

是否可以找到一种普遍的算法，用来判定一个任意的丢番图方程是否有整数解？

这样的问题在数学上被称为判定问题，因为它是寻求对数学问题进行判定的算法。



问题探索

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

希尔伯特在提出这个问题的时候，希望能直接找到这样的判定算法。

什么是算法？当时并没有一个明确的定义，这个困难使得这个问题在提出后的30多年内没有取得任何实质性的进展。直到20世纪30年代，人们对算法的研究才逐渐深入。

在数学上，算法是对数学函数进行有效计算的方法。寻找可以有效计算的函数是算法研究的一个重要切入点。



递归函数

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

一些最简单的函数，以及用这些函数通过简单规则组合出的函数。

1931年，赫尔布兰德(Herbrand)对递归函数进行了研究，并将结果寄给了著名的逻辑学家哥德尔(Gödel)，但没有得到回应。



λ -演算

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

丘奇(Church)也在从事逻辑与算法方面的研究，发展了一种逻辑体系。他的学生克林(Kleene)和罗瑟(Rosser)对这个体系做了细致的研究，发现这个体系竟然是自相矛盾的！但是其中有一部分是自治的，保留下来形成了 λ -演算。

λ -演算可以定义函数，而且所有用 λ -演算定义的函数都是可以有效计算的。



可以有效计算的函数

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

丘奇大胆猜测：所有可以有效计算的函数都可以由 λ -演算定义。

哥德尔以赫尔布兰德寄给他的递归函数为基础，对这以结果进行了完善，形成了赫尔布兰德-哥德尔递归函数。

丘奇和克林发现两种体系是等价的。



图灵(Turing)的研究

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

图灵研究什么样的运算可以用机器来实现。

图灵将他的工作与丘奇、哥德尔的工作做了比较，发现丘奇和哥德尔所定义的那些函数与他的抽象计算机可以计算的函数恰好吻合。



递归可枚举集

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

由可以有效计算的函数生成的自然数的集合。

人们发现一个微妙的结果：对于某些递归可枚举集来说，我们无法判定一个自然数是否属于该集合。



哥德尔不完全性定理

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

任何足够复杂并且自治的数学体系都必定包含不可判定问题。(1932年)

1947年，波斯特(Post)发现了第一个逻辑领域以外的不可判定问题。

人们意识到希尔伯特第十问题可能会有一个否定的回答。



丢番图集

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

1950年，戴维斯(Davis)引入了丢番图集的概念：这是一些由自然数组成的集合，是通过丢番图方程生成的。

丢番图集与递归可枚举集之间存在一种关联。如果存在算法判定丢番图方程有解，则可以判定一个自然数是否属于某个丢番图集。反之，如果证明了某些丢番图集是不可判定的，就可以证明希尔伯特第十问题有一个否定的回答。



有代价的解决方法

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

如果证明所有的递归可枚举集都是丢番图集，就证明了丢番图集是不可判定的，从而完成希尔伯特第十问题的否定解答。在戴维斯的证明中，用到了一个全称有界量词的逻辑符号。

为了去掉这个量词，戴维斯和普特南(Putnam)在9年后，通过引入两个假设，达到了这个目的。



罗宾逊(Robinson)猜想

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

罗宾逊夫人发现其中一个假设可以去掉，并简化了证明。但是对另一个假设，仍没有解决办法，即证明中用到了一种称为“指数丢番图集”的集合。它类似于丢番图集，但是涉及到了指数函数。

罗宾逊猜想：指数丢番图集实际上就是丢番图集。



问题解决

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

1969年，罗宾逊夫人为了证明罗宾逊猜想，提出了一个巧妙的思路。

马蒂亚塞维奇(Matiyasevich)看到了罗宾逊夫人的论文，立即被这个思路吸引。在1970年，他成功地证明了罗宾逊猜想，从而一举解决了希尔伯特第十问题。



自动机理论

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

自动机理论是研究离散数字系统的功能和结构以及两者关系的数学理论，是理论计算机科学的基础理论之一。离散数字系统的时间坐标是离散的，表征系统的变量的值是数字量。数字电路、数字信道、自动电话交换机、数字计算机、程序和算法是数字系统的一些实例。



源于人们对计算的探索

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究历史

课程中的概念

19世纪中期，布尔用数学方法研究思维规律的问题建立了逻辑代数，后人称之为布尔代数。1935至1938年，肖斯塔科夫和仙农，独立地应用布尔代数于继电器接点电路的分析和综合，形成了开关网络理论。1936年，为了对能行性、机械过程或算法进行精确的数学描述，图灵提出一种理想计算机，后人称之为图灵机。1944年，麦克卡洛和匹茨用数理逻辑方法研究神经网络。



源于人们对计算的探索

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

40年代中期出现电子计算机以后，美籍匈牙利数学家诺伊曼在1948年提出建立自动机的一般逻辑理论，对各种人造自动机和天然自动机进行比较性研究，探索其共同规律。他还研究了自动机的自繁殖和自恢复问题。诺伊曼被认为是自动机论的创立者。



成为独立的学科

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

受计算机的影响，50年代初，开关网络的研究重点转移到一般的逻辑网络，特别是门电路类型开关网络。1954年前后形成了有限自动机这一重要数学概念。同时，从数字计算机的一种理想的数学模型的角度，开始对图灵机深入研究。1956年《自动机研究》论文集的出版，标志着自动机论已形成为一门独立的学科。



自动机论研究的问题

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- ① 有限自动机论
- ② 无限自动机论
- ③ 概率自动机论
- ④ 细胞自动机论
- ⑤ 抽象自动机论



与其他学科的关系

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

- 自动机理论的形成和发展过程中，数理逻辑起了很大的作用。
- 自动机理论与形式语言理论、计算复杂性理论等关系密切。
- 自动机理论是控制论的一部分。



自动机理论的应用

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

- 自动机理论提供各种数字电路综合分析的系统方法；
- 作为一种基本工具用于高级程序语言的编译设计；
- 线性自动机可作为伪随机序列产生器；
- 可逆自动机可用于通信编码；
- 自动机也用于研究生物发育。



形式语言理论

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

用数学方法研究自然语言(如英语)和人工语言(如程序设计语言)的语法的理论。

形式语言只研究语言的组成规则，不研究语言的含义。形式语言理论在自然语言的理解和翻译、计算机语言的描述和编译、社会和自然现象的模拟、语法制导的模式识别等方面有广泛的应用。



历史情况

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

形式语言的研究始于20世纪初，用于模拟自然语言是在50年代。当时，数理语言学家用数学方法研究自然语言的结构。在电子计算机出现后，人们想到用计算机来作自然语言的机械翻译。在取得一些成功后便停滞不前，翻译质量很难提高，这因为对自然语言结构的理解太表面化。1956年，乔姆斯基发表了用形式语言方法研究自然语言的第一篇文章，给出了语言的一种定义方法。



历史情况

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历史

形式语言研究
历史

课程中的概念

1960年，算法语言ALGOL60报告发表。次年又发表了修改报告。在报告中，第一次使用一种称为BNF范式的形式方法来描述程序设计语言的语法。人们即发现BNF范式极其类似于形式语言理论中的上下文无关文法，从而打开了形式语言广泛应用于程序设计语言的局面，并给形式语言理论本身的研究以极大的推动，使它发展成为理论计算机科学的一个重要分支。



形式语言 and 文法

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

乔姆斯基用变换文法作为形式语言的描述手段。

变换文法定义成四元组 $G = (\Sigma, V, S, P)$ 。

根据产生式的结构。可以将文法分为零型文法、一型文法、二型文法和三型文法，分别生成零型语言、一型语言、二型语言和三型语言。



形式语言和自动机

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第十问题

自动机研究历史

形式语言研究历史

课程中的概念

乔姆斯基分层法提出不久，人们即发现它和自动机的分类有密切的关系。到1964年，四类文法及其语言全部在自动机中找到了它们所对应的语言。

零型、一型、二型和三型语言正好分别是图灵机、非确定型线性有界自动机、非确定型下推自动机和有限自动机接受的语言。



课程中的概念

形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

- 自动机(automata): 接受输入, 产生输出, 可能还有临时存储空间。
- 形式语言(formal language): 包括符号集和把符号组成句子的构成规则。
- 算法(algorithm): 把机械计算的概念形式化。



形式语言与
自动机理论

姚刚

目录

课程简介

希尔伯特的第
十问题

自动机研究历
史

形式语言研究
历史

课程中的概念

谢谢！

主 讲 人： 姚 刚

电子邮箱： yaogang@iie.ac.cn