形式语言与自动机理论
Formal Languages and Automata Theory

姓名：于萍

邮箱：**yuping0428@hit.edu.cn**

课程简介

- **学时：32学时（无实验）**
- **期末考试80%+大作业10%+平时成绩10%**

参考文献：

- 蒋宗礼，姜守旭. 形式语言与自动机理论(第4版), 清华大学出版社，2023
- John E Hopcroft. Introduction to Automata Theory, Languages, and Computation (2nd Edition). Addison-Wesley Publishing Company, 2001

课程简介

计算理论　　　　**计算理论**

数学　　　　**计算机**

- 计算机的<u>基本能力</u>和<u>限制</u>究竟是什么？

可计算性　　　　计算复杂性

1. 可计算理论

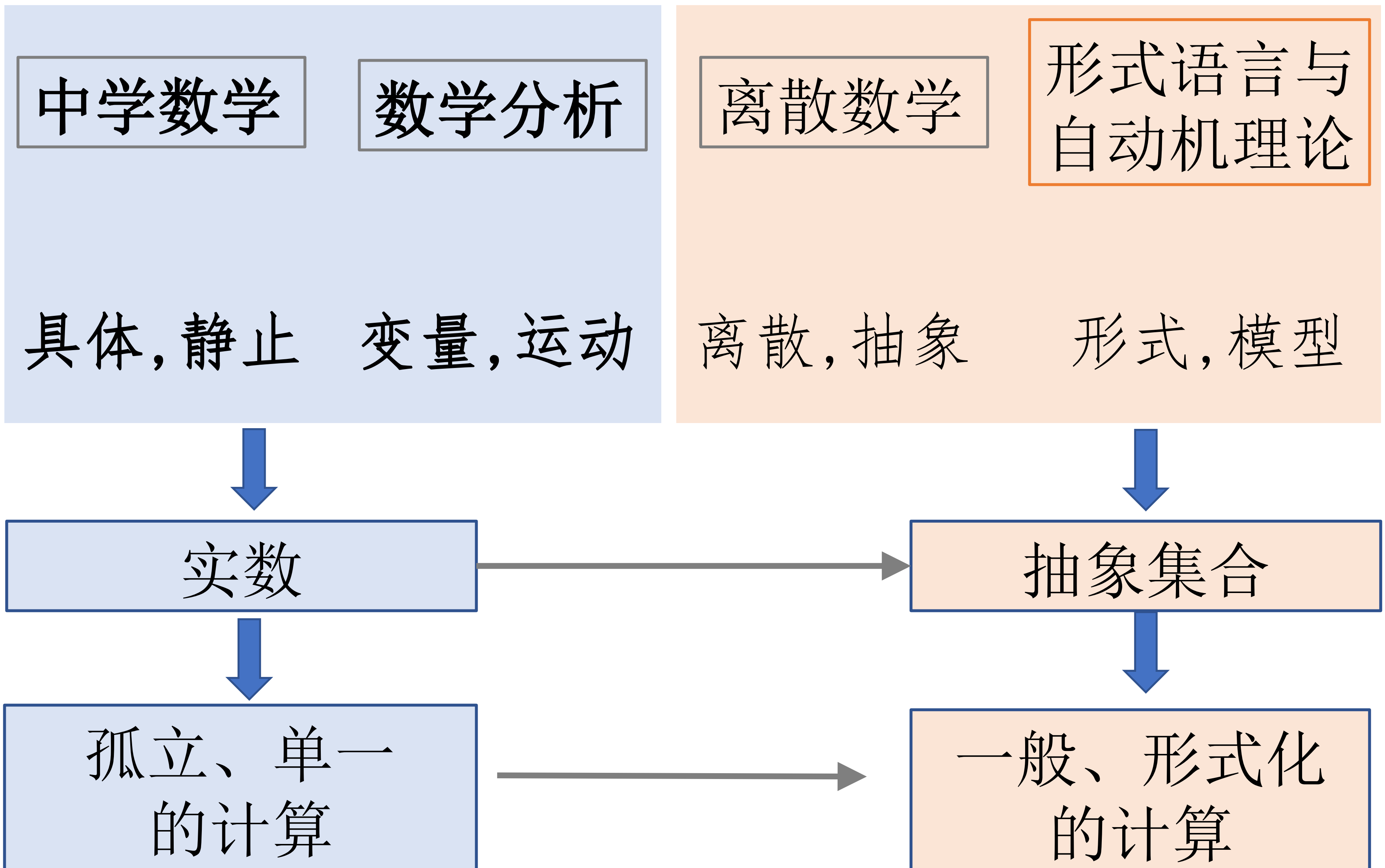
- 哪些问题可以通过计算来解决？
- 是不是任何问题都可以通过计算来解决？
- 如果不是，哪些可以，哪些不可以？为什么？

2. 计算复杂性理论

- 利用计算来解决可计算的问题，消耗哪些资源？
- 计算时间和存储空间是多少？
- 如果一个问题求解过程都需要相当多的资源，那其中的原因是什么？

计算理论的研究内容

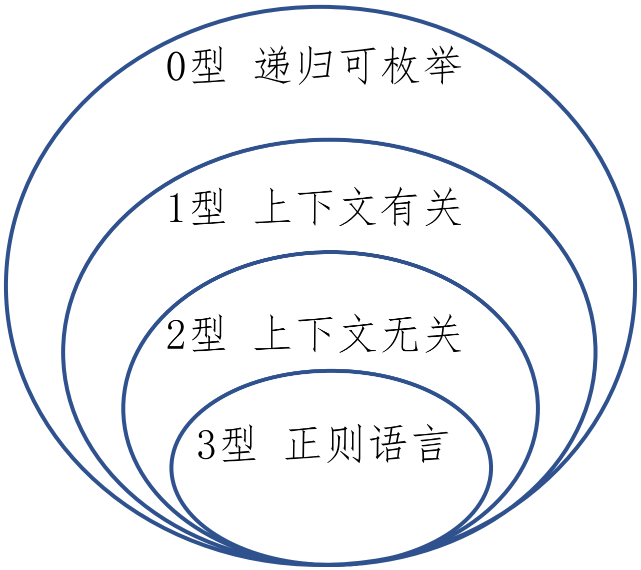
- 形式语言与自动机理论 (Automata Theory)
- 可计算性 (Computability Theory)
- 计算复杂性 (Complexity Theory)



中学数学
数学分析
离散数学
形式语言与
自动机理论
具体，静止
变量，运动
离散，抽象
形式，模型
实数
抽象集合
孤立、单一
的计算
一般、形式化
的计算

形式语言与自动机

- 1956年， Chomsky从语言产生的角度定义了语言与文法；
- 1951-1956年，Kleene提出了有穷状态自动机FA，从语言识别的角度定义了语言；
- 1959年，Chomsky证明了语言与自动机的等价性，形式语言从此诞生。



乔姆斯基文法体系

形式语言与自然语言的区别：形式语言只研究语法，不研究语义。

形式语言与自动机

计算模型/自动机：

FA	PDA	LBA	TM
Finite Automata	**Push Down Automata**	**Linear Bounded Automata**	**Turing Machine**
有穷状态自动机	下推自动机	线性有界自动机	图灵机

语言：

- 正则语言
- 3型语言

- 前后文无关语言
- 2型语言

- 前后文有关语言
- 1型语言

- 递归可枚举
- 0型语言

课程特点

- 培养学生“计算思维”能力，即问题的形式化描述、抽象思维能力和逻辑思维能力，同时了解和初步掌握“问题、形式化描述、自动化（计算机化）”的解题思路。
- 本课程的特点是抽象和形式化，既有严格的理论证明，又具有很强的构造性。
- 很难联系实际，但很重要。

课程特点

- 先修课程：

《高等数学》《离散数学》《集合论与图论》《数据结构》

- 后续课程：

《编译原理》

第一章 绪论

章节目录

1.1 集合的基础知识

集合的定义

■<u>集合：</u>一定范围内确定的，并且彼此可以区分的对象汇集在一起形成的整体叫做集合(set)，简称为集。

- 简单地说，集合是具有某种性质的<u>对象</u>的全体。

■<u>元素：</u>集合的成员为该集合的元素(element) 。

- 对象和集合的关系：该对象要么是集合中的一个元素，要么不是该集合的元素，二者必居其一。

1.1.1 集合及其表示

集合的表示方法

- **通常用大写的英文字母A, B, C, ... 和大写的希腊字母Γ, Σ, …表示集合;**
- **用小写字母a, b, c, …表示集合的元素。**

- 如果对象a是集合A中的一个元素，则记为$a \in A$，读作a属于A，或者A含有a；否则记为$a \notin A$，读作a不属于A，或者A不含a。

1.1.1 集合及其表示

集合的两种描述方法

1. **列举法：将所有的元素逐一地列举在大括号{}中，读者能立即看出规律时，某些元素可用省略号表示。**

例1-1 集合的列举表示。

① A = {1，3，6，9，10}。

② B = {本科生，硕士研究生，博士研究生}。

③ C = {0，5，10，15，… ，200}。

集合的两种描述方法

2. 命题法：基本形式为$\{x \mid P(x)\}$，其中P为谓词，此集合包括所有使P为真的x。

例1-2 集合的命题表示。

① $\mathrm{A} = \{x \mid 0 \le x \le 200\}$。

② $\mathrm{B} = \{x \mid 3x^2 + 8x + 4 = 0\}$。

集合的基数

<u>定义1-1</u> 如果集合A, B之间有一个一一对应，则称他们具有相同的基数，通常用|A|表示集合A的基数。

- 有穷集的基数就是它所包含元素的个数.

例1-3 有穷集合的基数。

① $\left|\{x|0 \leq x \leq 200 \text{ 且} \left(\exists \mathrm{n} \in \mathbf{N}\ (n \cdot 5 = x)\right)\}\right| = 41$

② $|\{a, b, c, \ldots, z\}| = 26$

1.1 集合的基础知识

1.1.1	集合及其表示
1.1.2	集合之间的关系
1.1.3	集合的运算

集合的关系：子集和真子集

定义1-2 设A, B是两个集合，如果集合A中的元素都是集合B的元素，则称集合A是集合B的子集(subset)，集合B是集合A的包集(container)。记作$A \subseteq B$，或$B \supseteq A$。

定义1-3 设A, B是两个集合，如果$A \subseteq B$，且$\exists x \in B$，但$x \notin A$，则称A是B的真子集(proper subset)，记作$A \subset B$。

集合的关系：相等

定义1-4 如果集合A, B含有的元素完全相同，则称集合A与集合B相等，记作$A = B$。

例1-4 集合的关系。

① 如果$A \subseteq B$且$B \subseteq A$，则A=B。

② 如果$A = B$，则$|A| = |B|$。

③ 如果A是有穷集，且$A \subset B$，则$|A| < |B|$。

④ 如果$A \subset B$，则$|A| \leq |B|$。

1.1 集合的基础知识

- ■并
- ■交
- ■差
- ■笛卡尔积
- ■幂集
- ■补集

集合的运算：并

定义1-5 设A, B是两个集合，A与B的并是一个集合，该集合中的元素要么是A的元素，要么是B的元素，记作$A \cup B$。

$$A \cup B = \{a \mid a \in A \text{或} a \in B\}。$$

定义1-5' 设$A_1, A_2, \cdots, A_n$是n个集合，则他们的并记作$\bigcup_{i=1}^{n} A_i$。

$$A_1 \cup A_2 \cup \cdots \cup A_n = \{a \mid \exists i, 1 \le i \le n, \text{使得} a \in A_i\}。$$

集合的运算：交

定义1-6 设A, B是两个集合，A与B的交是一个集合，该集合是由既属于A，又属于B的元素组成，记作$A\cap B$。

$$A\cap B=\{a\mid a\in A且a\in B\}。$$

- 如果$A\cap B=\emptyset$，则称A与B不相交。

集合的运算：差

定义1-7 设A, B是两个集合，A与B的差是一个集合，该集合是由属于A，但不属于B的所有元素组成，记作$A-B$。

$$A-B=\{a\,|a\in A\text{且}a\notin B\}。$$

例1-5 集合的差。

① 设A={1, 3, 5, …}，B ={2, 4, 6, …}，则$A-B=$ A。

② 设A={1, 2, 3, 4}，B ={2, 4, 6, …}，则$A-B=$ {1, 3}。

集合的运算：对称差

定义1-8 设A, B是两个集合，A与B的**对称差**是一个集合，该集合由属于A但不属于B，以及属于B但不属于A的所有元素组成。

- 记作$A \oplus B$，读作A对称减B。

$$A \oplus B = \{a \mid a \in A 且 a \notin B 或 a \notin A 且 a \in B\}。$$

- 显然，对集合A与B，有

$$A \oplus B = (A - B) \cup (B - A) = (A \cup B) - (A \cap B)$$

集合的运算：笛卡尔积

定义1-9 设A, B是两个集合，A与B的笛卡尔积是一个集合，该集合由所有这样的有序对(a,b)组成，其中$a \in A, b \in B$，记作$A \times B$。

$$A \times B = \{(a,b) \mid a \in A \text{且} b \in B\}。$$

例1-6 集合的笛卡尔积。

设A={1, 3, 5, …}，B ={2, 4, 6, …}，则$A \times B =$ $\{(a,b) \mid a$是任意的正奇数，b是任意的正偶数$\}$。

集合的运算：幂集

定义1-10 设A是一个集合，A的幂集是一个集合，该集合由A的所有子集组成，记作2^A，读作A的幂集。

$$2^A = \{B|B \subseteq A\}。$$

例1-7 集合的幂集。

设A={1, 2, 3}，则

$$2^A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}。$$

集合的运算：补集

定义1-11 设A是论域U上的一个集合，A的补集是一个集合，该集合由在U中，但不在A中的所有元素组成，记作$\bar{A}$，读作A（关于论域U）的补集。

$$\bar{A} = \mathrm{U} - \mathrm{A}。$$

例1-8 集合的补集。

设U={1, 2, 3, 4, 5}， A={4, 2}则 $\bar{A} = \{1, 3, 5\}$。

章节目录

1.2 关系

二元关系

集合{1, 3, 4, 8}和集合{0, 3, 5, 7}的元素之间存在的小于关系有：

1<3， 1<5， 1<7， 3<5， 3<7， 4<5， 4<7

(1,3)，(1,5)，(1,7)，(3,5)，(3,7)，(4,5)，(4,7)

存在的大于关系有:

1>0， 3>0， 4>0， 4>3， 8>0， 8>3， 8>5， 8>7

(1,0)，(3,0)，(4,0)，(4,3)，(8,0)，(8,3)，(8,5)，(8,7)

1.2.1 二元关系

二元关系的定义

<u>定义1-12</u> 设A, B是两个集合，任意的$R \subseteq (A \times B)$，R是A到B的二元关系。

- (a, b)$\in R$表示a与b满足关系R，也可表示为aRb。
- A称为定义域，B称为值域。
- 当A=B时，则称R是A上的二元关系。

二元关系的性质

定义1-13 设R是A上的二元关系，

① 如果对任意的a∈A，有(a, a)∈R，则称R是自反的。

② 如果对任意的a∈A，有(a, a) ∉ R，则称R是反自反的。

③ 如果对任意的a, b∈A，当(a, b) ∈ R时，必有(b, a) ∈ R，则称R是对称的。

④ 如果对任意的a, b∈A，当(a, b)∈R和 (b, a)∈R同时成立时，必有a=b，则称R是反对称的。

⑤ 如果对任意的a, b, c∈A，当(a, b)∈R和(b, c)∈R同时成立时，必有(a, c)∈R，则称R是传递的。

二元关系的性质

关系的三歧性：
自反性、对称性、传递性

例1-9 关系的性质。

- “=”关系是自反的、对称的、传递的；
- “<”“>”关系是反自反的、传递的；
- “≤”“≥”关系是自反的、反对称的、传递的。

- 通常意义下的父子关系是？

反自反的

1.2 关系

- 1.2.1 二元关系
- 1.2.2 等价关系与等价类
- 1.2.3 关系的合成
- 1.2.4 递归定义与归纳证明
- 1.2.5 关系的闭包

等价关系

<u>定义1-14</u> 如果集合A上的二元关系R是自反的、对称的、传递的，则称R是等价关系。

Callback:

- 如果对任意的$a \in A$，有$(a,a) \in R$，则称R是自反的。
- 如果对任意的$a,b \in A$，当$(a,b) \in R$时，必有$(b,a) \in R$，则称R是对称的。
- 如果对任意的$a,b,c \in A$，当$(a,b) \in R$和$(b,c) \in R$同时成立时，必有$(a,c) \in R$，则称R是传递的。

等价类

定义1-15 设R是集合S上的等价关系，则满足如下要求的S的划分$S_1, S_2, S_3, \cdots, S_n, \cdots$称为**S关于R的等价划分**，$S_i$称为**等价类**。

- $S = S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_n \cup \cdots$;
- 任意的$i \neq j$，有$S_i \cap S_j = \emptyset$；
- 任意的i，集合S_i中的任意两个元素a和b，aRb恒成立;
- 任意的$i \neq j$，S_i中的任意元素a和S_j中的任意元素b，aRb恒不成立。

等价关系在集合上的指数

R将S分成的等价类的个数称为R在S上的指数。

- **有时R可将S分成有穷多个等价类，此时称R具有有穷指数；**
- **有时R可将S分成无穷多个等价类，此时称R具有无穷指数。**

1.2.2 等价关系与等价类

等价类

例1-10 非负整数集上的模5同余关系将N={0, 1, 2, 3, …}

分成5个等价类:

- S_0={0, 5, 10, 15, 20, …}
- S_1={1, 6, 11, 16, 21, …}
- S_2={2, 7, 12, 17, 22, …}
- S_3={3, 8, 13, 18, 23, …}
- S_4={4, 9, 14, 19, 24, …}

Callback:

1. N= $S_0 \cup S_1 \cup S_2 \cup S_3 \cup S_4$
2. $S_i \cap S_j = \emptyset$;
3. S_i中任意a和b，aRb恒成立
4. S_i中的任意元素a和S_j中的任意元素b，aRb恒不成立。

1.2 关系

1.2.3 关系的合成

定义1-16 设$\boldsymbol{R_1 \subseteq A \times B}$是$\boldsymbol{A}$到$\boldsymbol{B}$的关系，$\boldsymbol{R_2 \subseteq B \times C}$是$\boldsymbol{B}$到$\boldsymbol{C}$的关系，则$\boldsymbol{R_1}$与$\boldsymbol{R_2}$的合成$\boldsymbol{R_1 \circ R_2}$是$\boldsymbol{A}$到$\boldsymbol{C}$的关系。

$$\boldsymbol{R_1 \circ R_2 = \{(a,c) \mid \exists (a,b) \in R_1 \text{ 且 } (b,c) \in R_2\}}$$

- 关系的合成运算符“∘”可以省略不写，如$R_1 \circ R_2$可以写成R_1R_2。

1.2.3 关系的合成

例1-11 设R_1和R_2是集合{1, 2, 3, 4}上的关系，其中

R_1={(1, 1), (1, 2), (2, 3), (3, 4)}

R_2={(2, 4), (4, 1), (4, 3), (3, 1), (3, 4)}

则

$R_1 \circ R_2$ = {(1, 4), (2, 1), (2, 4), (3, 1), (3, 3)}。

关系的合成的性质

设$\boldsymbol{R_1}, \boldsymbol{R_2}, \boldsymbol{R_3}$分别是S上的二元关系，可以证明以下结论：

- $\boldsymbol{R_1 R_2 \neq R_2 R_1}$
- $\boldsymbol{(R_1 R_2) R_3 = R_1 (R_2 R_3)}$ （结合律）
- $\boldsymbol{(R_1 \cup R_2) R_3 = R_1 R_3 \cup R_2 R_3}$ （合成对∪的右分配律）
- $\boldsymbol{R_3 (R_1 \cup R_2) = R_3 R_1 \cup R_3 R_2}$ （合成对∪的左分配律）
- $\boldsymbol{(R_1 \cap R_2) R_3 \subseteq R_1 R_3 \cap R_2 R_3}$ （合成对∩的右分配律）
- $\boldsymbol{R_3 (R_1 \cap R_2) \subseteq R_3 R_1 \cap R_3 R_2}$ （合成对∩的左分配律）

1.2 关系

递归定义的组成

递归定义可以用来定义一个集合。集合的递归定义由三部分组成：

- 基础：用来定义该集合的最基本的元素。
- 归纳：指出用集合中的元素来构造集合的新元素的规则。
- 极小性限定：指出一个对象是所定义集合中的元素的充要条件是它可以通过有限次的使用基础和归纳条款中所给的规定构造出来。

1.2.4 递归定义与归纳证明

例1-12 著名的斐波那契数的定义：

⑴ 基础：0是第一个斐波那契数，1是第二个斐波那契数。

⑵ 归纳：如果n是第i个斐波那契数，m是第i+1个斐波那契数，则n+m是第i+2个斐波那契数，其中i是大于等于1的正整数。

(3) 只有满足(1)和(2)的数才是斐波那契数。

递归定义使得集合中元素的**构造规律**明确地表现出来，给集合性质的归纳证明提供了良好基础。

归纳证明的组成

归纳证明与递归定义相对应，由三步组成：

- 基础：证明最基本元素具有相应性质。
- 归纳：证明如果某些元素具有相应性质，则根据这些元素用所规定的方法得到的新元素也具有相应的性质。
- 根据归纳法原理，集合中所有的元素具有性质P，即集合具有性质P。

1.2 关系

- 1.2.1 二元关系
- 1.2.2 等价关系与等价类
- 1.2.3 关系的合成
- 1.2.4 递归定义与归纳证明
- **1.2.5 关系的闭包**

1.2.5 关系的闭包

定义1-17 设R是S上的二元关系，R的**正闭包R^+**定义为：

(1) $R \subseteq R^+$；

(2) 如果(a, b), (b, c) $\in R^+$，则(a, c) $\in R^+$；

(3) 除(1), (2)外，R^+不再含有其他任何元素。

对任意二元关系**R**，有

$$\boldsymbol{R}^+ = \boldsymbol{R} \cup \boldsymbol{R}^2 \cup \boldsymbol{R}^3 \cup \boldsymbol{R}^4 \cup \cdots$$

$$= \boldsymbol{R} \cup \boldsymbol{R}^2 \cup \boldsymbol{R}^3 \cup \cdots \cup \boldsymbol{R}^{|S|},$$ 当S为有穷集时

<u>定义1-18</u> 设R是S上的二元关系，R的**克林闭包** R^*定义为：

(1) $R^0 \subseteq R^*, R \subseteq R^*$;

(2) 如果(a, b), (b, c) ∈ R^*， 则(a, c) ∈ R^* ；

(3) 除(1), (2)外，R^*不再含有其他任何元素。

$$R^* = R^0 \cup R^+$$
$$= R^0 \cup R \cup R^2 \cup R^3 \cup R^4 \cup \cdots$$
$$= R^0 \cup R \cup R^2 \cup R^3 \cup \cdots \cup R^{|S|}\text{，} \quad \text{当S为有穷集时}$$

章节目录

1. 3 图

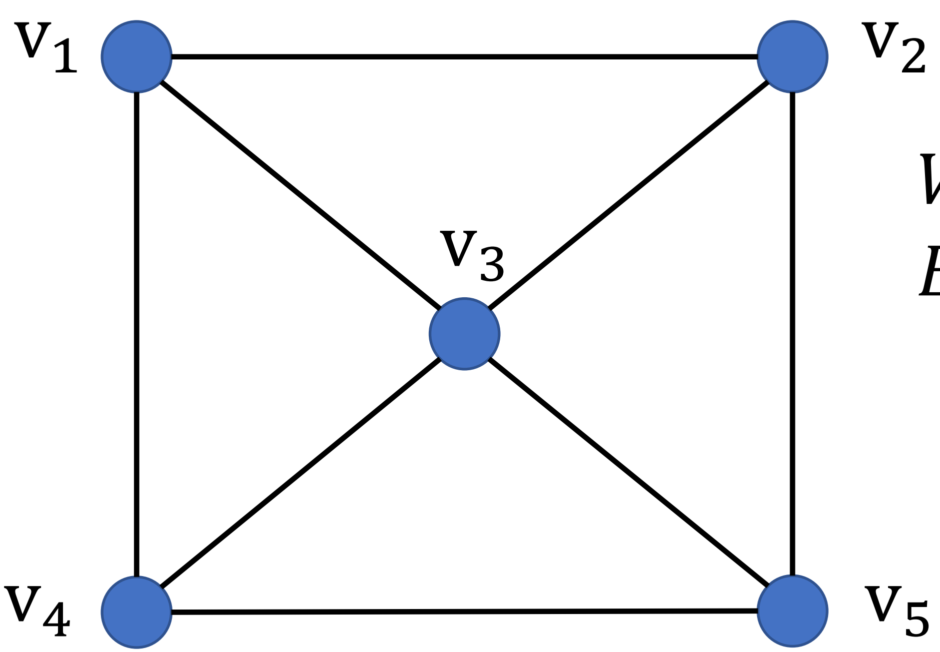
1.3.1 无向图

1.3.2 有向图

1.3.3 树

1.3.1 无向图

定义1-19 $G=(V,E)$表示一个无向图，V是一个非空有穷集合，V中的元素称为顶点，$E \subseteq V \times V$，E中的元素称为无向边。



$$V=\{v_1, v_2, v_3, v_4, v_5\}$$

$$E=\{(v_1,v_2),(v_1,v_3),(v_1,v_4),(v_2,v_3),(v_2,v_5),(v_3,v_4),(v_3,v_5),(v_4,v_5)\}$$

1.3.1 无向图

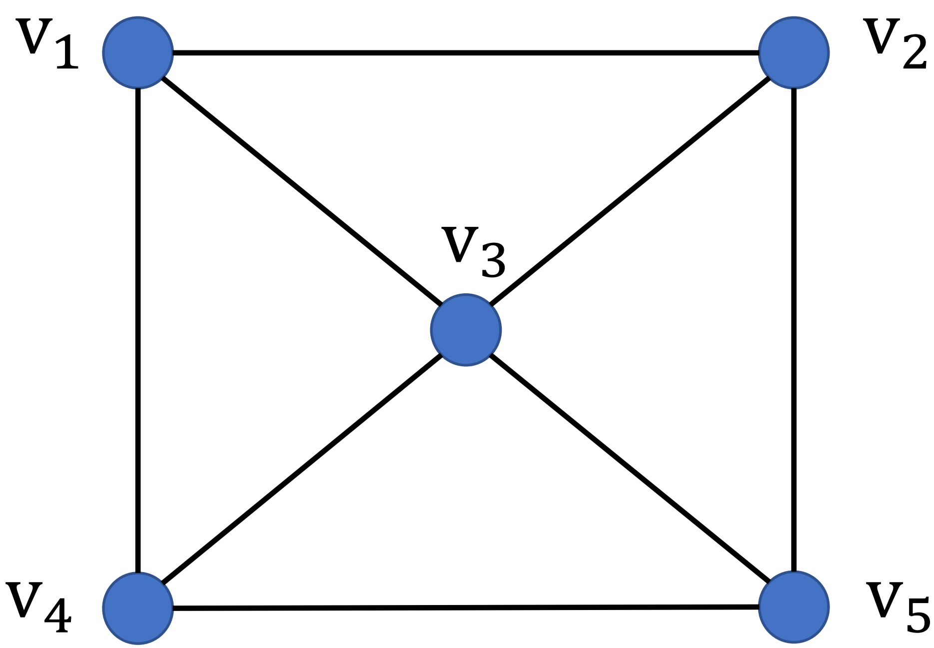
回路、顶点的度数

定义1-20 设$G=(V,E)$是一个无向图，如果对于$0\le i\le k-1\ (k\ge 1)$均有$(v_i,v_{i+1})\in E$，则称$v_0,v_1,\cdots,v_k$是G的一条长为k的路。当$v_0=v_k$时，$v_0,v_1,\cdots,v_k$是叫做一个回路或圈。

定义1-21 设$G=(V,E)$是一个无向图，对于$v\in V$, $|\{w|(v,w)\in E\}|$ 称为顶点v 的度数。

连通图

定义1-22 设$G=(V,E)$是一个无向图，若对于$\forall v,w\in V, v\neq w$，$v$与$w$之间至少有一条路存在，则称G是连通图。



1. 3 图

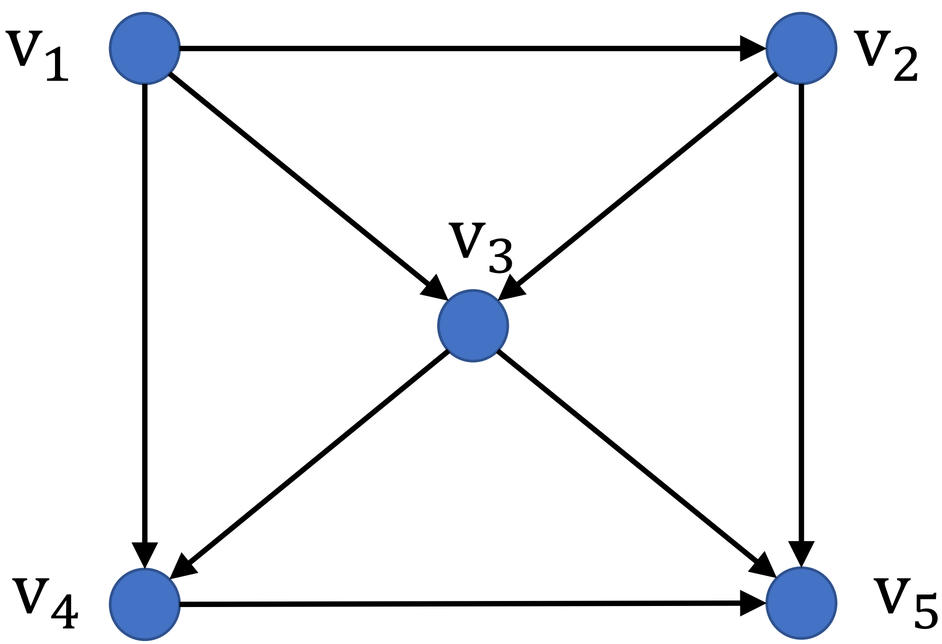
1.3.1 无向图

1.3.2 有向图

1.3.3 树

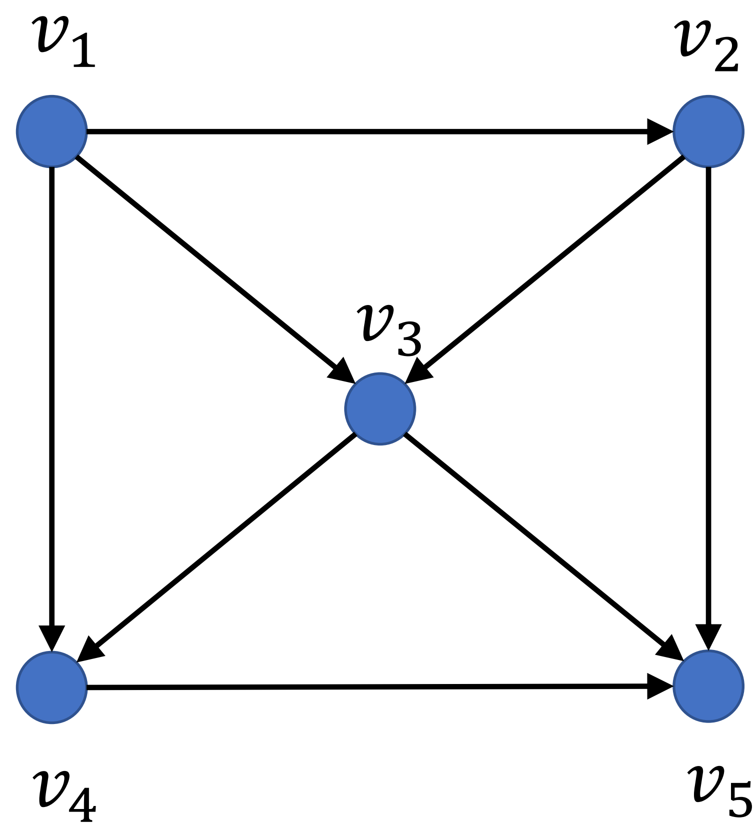
1.3.2 有向图

定义1-23 设V是一个非空有穷集合，$E \subseteq V \times V$，称$G = (V, E)$为一个有向图，其中V中的元素称为顶点，$\forall (v_1, v_2) \in E$ 称为从顶点v_1到v_2的有向边，E称为有向边集。



1.3.2 有向图

顶点的入度和出度



定义1-24 设$G = (V, E)$是一个有向图，对$\forall v \in V$，

$\mathrm{ideg}(v) = |\{w | (w, v) \in E\}|$

$\mathrm{odeg}(v) = |\{w | (v, w) \in E\}|$

其中， $\mathrm{ideg}(v)$ 称为顶点v的入度，$\mathrm{odeg}(v)$称为顶点v的出度。

1.3 图

1.3.3 树

前导：某有向边如右图，称v_1为v_2的前导

定义1-25 设$G=(V,E)$是一个有向图，当G满足如下条件时，称G为一颗树：

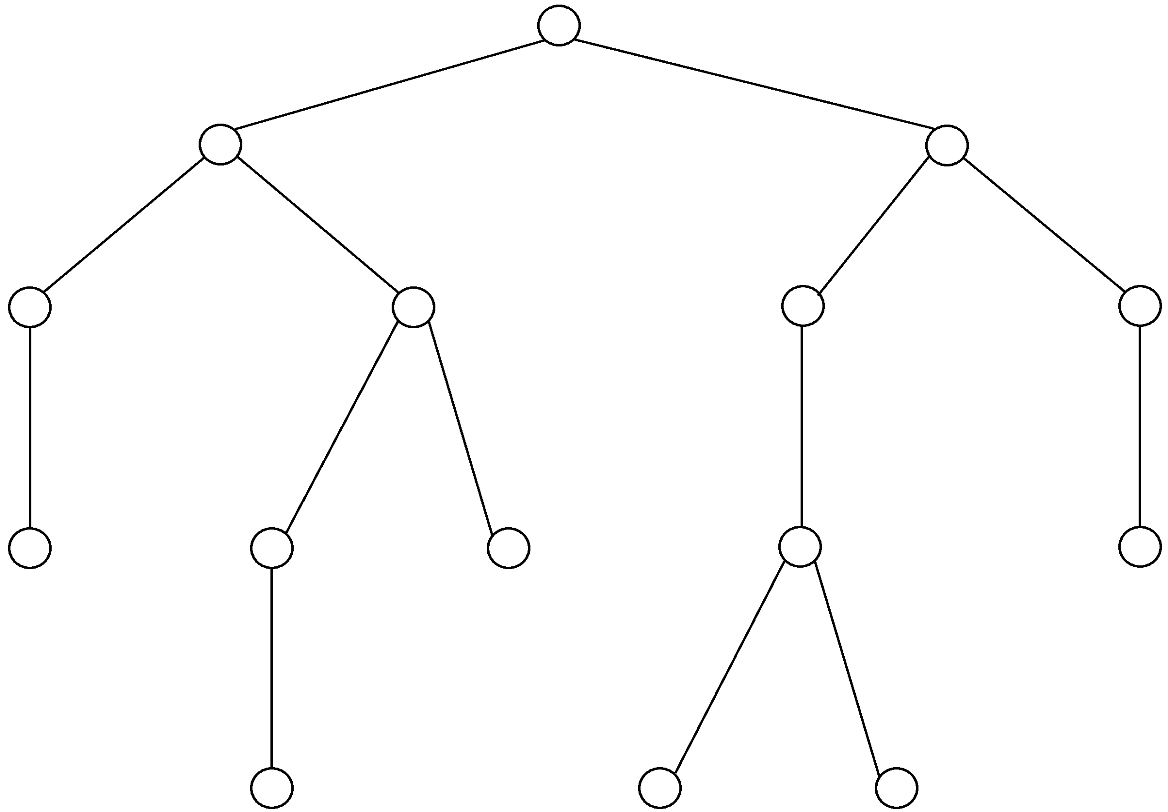
(1) $\exists v \in V$, v没有前导，且v到树中其他顶点均有一条有向路，称此顶点为树G的根；

(2) 每个非根顶点有且仅有一个前导；

(3) 每个顶点的后继按其拓扑关系从左到右排序。

1.3.3 树

定义1-26 设$G=(V,E)$是一颗树，如果对于$\forall v\in V$, v最多只有两个儿子，则称G为二元树。



章节目录

1.4 语言

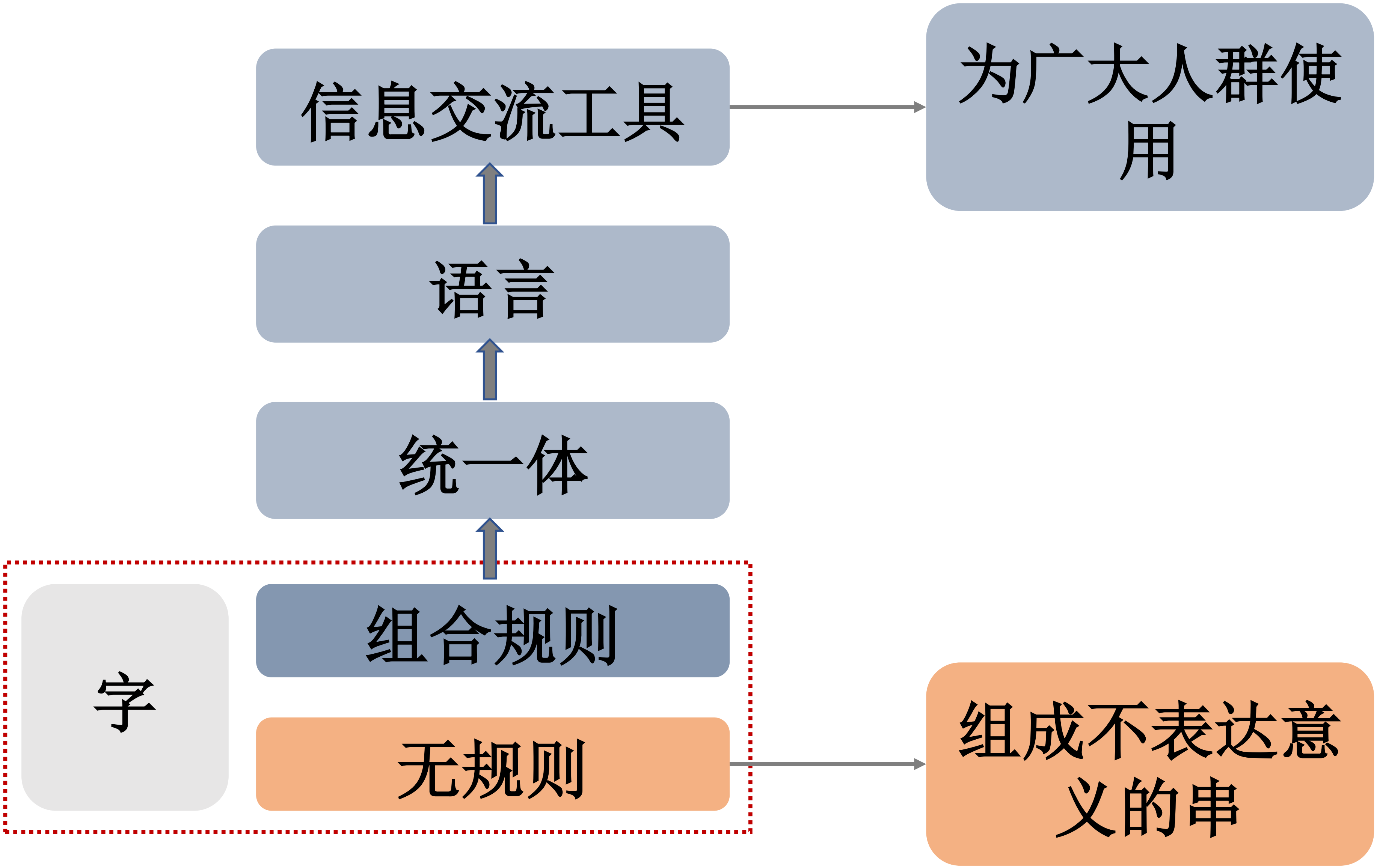
- 1.4.1 什么是语言
- 1.4.2 形式语言与自动机理论的产生与作用
- 1.4.3 基本概念

1.4.1 什么是语言

语言是一定的群体用来进行交流的工具，必须有着一系列的生成规则、理解(语义)规则，只有当使用者按照这些规则来构造“句子”和理解“句子”时，才能达到交流的目的。

- “学大一生是个我”；“我是一个大学生”。

1.4.1 什么是语言



1.4.1 什么是语言

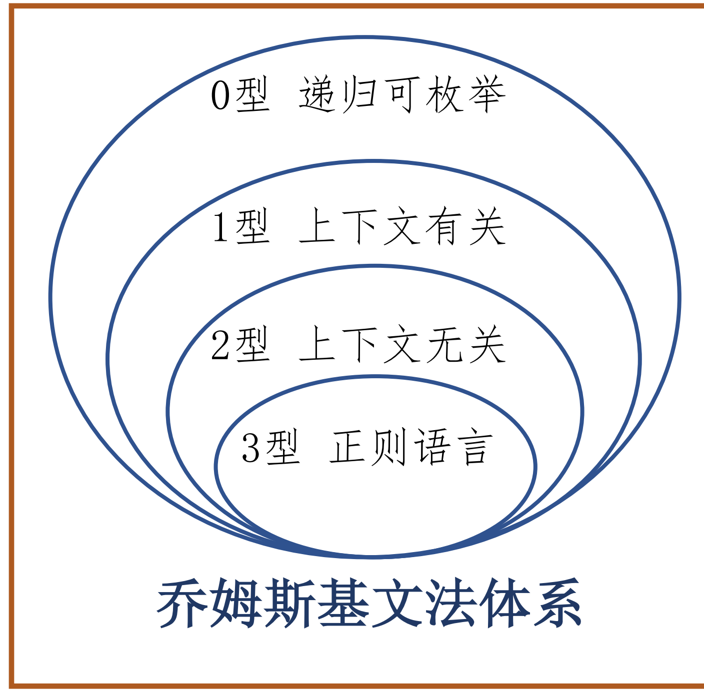
- 斯大林：从强调语言的作用出发，把语言定义为“为广大的人群所理解的字和组合这些字的方法”。
- 语言学家韦波斯特(Webster)：为相当大的团体的人所懂得并使用的字和组合这些字的方法的统一体。

用这些定义来建立语言的数学模型是不够精确的，必须有更形式化的定义。

1.4 语言

1.4.2 形式语言与自动机理论的产生与作用

- 1956年， Chomsky从语言产生的角度定义了语言与文法；
- 1951-1956年，Kleene提出了有穷状态自动机FA，从语言识别的角度定义了语言；



- 1959年，Chomsky证明了语言与自动机的等价性，形式语言从此诞生，并得到广泛研究。尤其是上下文无关语言被作为计算机程序设计语言的最佳近似得到了较为深入的研究。

1.4 语言

1.4.3 基本概念

定义1-27 字母表是一个非空有穷集合，字母表中的元素称为该字母表的一个字母。又叫做符号、或者字符。

字母表具有非空性和有穷性，通常用Σ表示。

- **{a，b，c，d}**
- **{a，b，c，…，z}**
- **{0, 1}**

定义1-28 设Σ_1, Σ_2是两个字母表， Σ_1与Σ_2的乘积：

$$\Sigma_1\Sigma_2 = \{ab|a \in \Sigma_1, b \in \Sigma_2\}$$

例1-13 字母表的乘积

(1) {0,1}{0,1} = {00, 01, 10, 11}。

(2) {0,1}{a,b,c,d} = {0a, 0b, 0c, 0d, 1a, 1b, 1c, 1d}。

(3) {aa, ab, bb}{0,1} = {aa0, aa1, ab0, ab1, bb0, bb1}。

1.4.3 基本概念

定义1-29 设Σ是一个字母表， Σ的n次幂递归地定义为:

(1) $\Sigma^{0} = \{\varepsilon\}$

(2) $\Sigma^{n} = \Sigma^{n-1}\Sigma, n \geq 1$

其中，ε是由Σ中的0个字符构成的空串。

1.4.3 基本概念

定义1-30 设Σ是一个字母表，Σ的正闭包：

$$\Sigma^{+} = \Sigma \cup \Sigma^{2} \cup \Sigma^{3} \cup \cdots$$

Σ的克林闭包：

$$\Sigma^{*} = \Sigma^{0} \cup \Sigma^{+} = \Sigma^{0} \cup \Sigma \cup \Sigma^{2} \cup \Sigma^{3} \cup \cdots$$

- $\Sigma^{+}=\{x \mid x$ 是由Σ中的至少一个字符连接而成的字符串$\}$；
- $\Sigma^{*}=\{x \mid x$ 是由Σ中的若干（大于等于0）个字符连接而成的字符串$\}$。
- Σ^{+}和Σ^{*}都是集合。

1.4.3 基本概念

- 句子：$\boldsymbol{\Sigma}$是一个字母表，$\forall \boldsymbol{x} \in \boldsymbol{\Sigma}^*$，$\boldsymbol{x}$叫做$\boldsymbol{\Sigma}$上的一个句子/字符串。
- 句子的长度：

• $\forall \boldsymbol{x} \in \boldsymbol{\Sigma}^*$, 句子$\boldsymbol{x}$中字符出现的总个数叫做该句子的长度，记作$|\boldsymbol{x}|$。

• 长度为0的字符串叫空句子，记作 $\boldsymbol{\varepsilon}$，$|\boldsymbol{\varepsilon}| = \mathbf{0}$。

思考：$\{\boldsymbol{\varepsilon}\}$与$\emptyset$?

• $\{\boldsymbol{\varepsilon}\} \neq \emptyset$
• $\boldsymbol{\varepsilon}$是一个句子，长度为0，而$\emptyset$是空集
• $|\{\boldsymbol{\varepsilon}\}|$=1，$|\emptyset| = \mathbf{0}$

1.4.3 基本概念

- 并置/连接：$\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\Sigma}^*$，$\boldsymbol{x}, \boldsymbol{y}$的并置/连接是由串$\boldsymbol{x}$直接与串$\boldsymbol{y}$相连组成的，记作$\boldsymbol{xy}$。
- 串$\boldsymbol{x}$的$\boldsymbol{n}$次幂
  - $\boldsymbol{x^0 = \varepsilon}$
  - $\boldsymbol{x^n = x^{n-1}x}$

例1-14 串的连接：串$x = 001$，$y = 1101$，则

(1) $xy = 0011101$；

(2) $x^0 = y^0 = \varepsilon$；

(3) $x^3 = 001001001$；

(4) $y^4 = 1101110111011101$。

1.4.3 基本概念

- 前缀与后缀

设$x, y, z \in \Sigma^*$，且$z = xy$，则

- x是z的前缀；

  如果$y \neq \varepsilon$，则x是z的真前缀；（$x \neq z$）
- y是z的后缀；

  如果$x \neq \varepsilon$，则y是z的真后缀。（$y \neq z$）

1.4.3 基本概念

■ 公共前缀与公共后缀

设$x, y, z, w, v \in \Sigma^*$,

- 若$w = xy, v = xz$，则称x是w和v的公共前缀；如果w和v的任何公共前缀都是x的前缀，则称x是w和v的最大公共前缀。
- 若$w = xz, v = yz$，则称z是w和v的公共后缀；如果w和v的任何公共后缀都是z的后缀，则称z是w和v的最大公共后缀。

例1-15 前缀和后缀。

字母表$\boldsymbol{\Sigma = \{a, b\}}$上的句子**abaabb**的前缀、真前缀、后缀和真后缀如下：

- 前缀： **ε, a, ab, aba, abaa, abaab, abaabb**
- 真前缀： **ε, a, ab, aba, abaa, abaab**
- 后缀： **ε, b, bb, abb, aabb, baabb, abaabb**
- 真后缀： **ε, b, bb, abb, aabb, baabb**

1.4.3 基本概念

■ 子串和公共子串

- 设$x, y, z, w \in \Sigma^*$，且$w = xyz$，称y是w的子串；
- 设$t, u, v, w, x, y, z \in \Sigma^*$，且$w = xyz$，$v = uyt$，则称$y$是$w$和$v$的公共子串；
- 若$y_1, y_2, \cdots, y_n$是w和v的公共子串，且$\max\{|y_1|, |y_2|, \cdots, |y_n|\} = |y_i|$，则称$y_i$是$t$和$w$的最大公共子串。

显然，两个串的最大公共子串并不一定是唯一的。

1.4.3 基本概念

定义1-31 $\forall L \subseteq \Sigma^*$，$L$称为字母表Σ上的一个**语言**。$\forall x \in L$，$x$叫做$L$上的一个句子。

例1-16 {0, 1}上的不同语言。

- {00, 11}，{0, 1}
- {0, 1, 00, 11}，{0, 1, 00, 11, 01, 10}
- $\{00, 11\}^*$，$\{01, 10\}^*$，$\{00, 01, 10, 11\}^*$
- $\{0\}\{0, 1\}^*\{1\}$，$\{0, 1\}^*\{111\}\{0, 1\}^*$，

显然，语言是一个集合。

1.4.3 基本概念

<u>定义1-32</u> $L_1 \subseteq \Sigma_1{}^*$， $L_2 \subseteq \Sigma_2{}^*$，语言$L_1$与$L_2$的乘积是一个语言，该语言定义为：

$L_1L_2 = \{xy|x \in L_1, y \in L_2\}$是字母表$\Sigma_1 \cup \Sigma_2$上的语言。

<u>定义1-33</u> $\forall L \subseteq \Sigma^*$， L的n次幂是一个语言，该语言被定义为：

(1) 当$n = 0$时， $L^n = L^0 = \{\varepsilon\}$

(2) 当$n \geq 1$时， $L^n = L^{n-1}L$

$L^+ = L \cup L^2 \cup L^3 \cup \cdots$ $\quad L^* = L^0 \cup L \cup L^2 \cup L^3 \cup \cdots$

章节目录

1.5 本章小结

1. 集合：集合的表示、集合之间的关系、集合的基本运算；
2. 关系：主要介绍了二元关系相关的内容。包括等价关系、等价分类、关系合成、关系闭包，递归定义与归纳证明；
3. 图：无向图、有向图、树的基本概念；
4. 语言与形式语言：自然语言的描述，形式语言和自动机理论的出现，字母表，字母表上的语言、语言的基本运算等。

Thanks!