11.串

(f1) Karp-Rabin算法：串即是数

All things are numbers.

- Pythagoras (570 ~ 495 BC)

God made the integers;

all else is the work of man.

- L. Kronecker (1823 ~ 1891)

邓俊辉

deng@tsinghua.edu.cn

凡物皆数

❖ 
Kurt Gödel 1906-1978

$\boxed{\text{Gödel numbering}}$

逻辑系统的符号、表达式、公式、命题、定理、公理等

均可以不同的 $\boxed{\text{自然数}}$ 标识

❖ 素数序列： $p(k)$ = 第k个素数

2 , 3 , 5 , 7 , 11 , ...

❖ K. Gödel： 每个有限维的自然数 $\boxed{\text{向量}}$ ，唯一对应于一个 $\boxed{\text{自然数}}$

$$\langle a_1, a_2, \ldots, a_n \rangle \sim p(1)^{1+a1} \times p(2)^{1+a2} \times \ldots \times p(n)^{1+an}$$

$$\langle \boxed{3}, \boxed{1}, \boxed{4}, \boxed{1}, \boxed{5}, \boxed{9}, \boxed{2}, \boxed{6} \rangle$$

$$= 2\wedge\boxed{4} \times 3\wedge\boxed{2} \times 5\wedge\boxed{5} \times 7\wedge\boxed{2} \times 11\wedge\boxed{6} \times 13\wedge\boxed{10} \times 17\wedge\boxed{3} \times 19\wedge\boxed{7}$$

凡物皆数

- K. Gödel：给定$\boxed{\text{可数}}$字母表，有限长度的字符串均$\boxed{\text{唯一}}$对应于自然数

$$\{ a , b , c , \ldots , z , \ldots \}$$

$$= \{ 1 , 2 , 3 , \ldots , 26 , \ldots \}$$

$$\boxed{\text{godel}} = \boxed{2^{1+7}} \times \boxed{3^{1+15}} \times \boxed{5^{1+4}} \times \boxed{7^{1+5}} \times \boxed{11^{1+12}}$$

$$= \boxed{139\ 869\ 560\ 310\ 664\ 817\ 087\ 943\ 919\ 200\ 000}$$

- 若果真如RAM模型所假设，$\boxed{\text{字长无限}}$

  则$\boxed{\text{只需一个}}$寄存器即可...

凡物皆数

- 反过来，由Gödel编号可否$\boxed{\text{唯一确定}}$原字符串？
- Book IX of *The Elements of Geometry* (ca 300 B.C.)

  Euclid: every number factors uniquely into primes
- 因此，由合法的编号，经素因子$\boxed{\text{分解}}$再$\boxed{\text{排序}}$，即可$\boxed{\text{唯一确定}}$原字符串
- 素因子分解，至今$\boxed{\text{尚无有效}}$算法！

  ——不幸？幸运？

凡物皆数

❖ Cantor numbering


Georg Cantor
(1845-1918)

❖ $cantor_2(\ \boxed{i},\ \boxed{j}\ ) = (\ (\ i + j\ )^2 + 3*i + j\ )\ /\ 2$

$cantor_2(\ \boxed{2},\ \boxed{3}\ ) = (\ (\ 2 + 3\ )^2 + 3*2 + 3\ )\ /\ 2 = 17$

$cantor_2(\ \boxed{3},\ \boxed{2}\ ) = (\ (\ 3 + 2\ )^2 + 3*3 + 2\ )\ /\ 2 = 18$

0	1	3	6	10	15
2	4	7	11	16	
5	8	12	17		
9	13	18			
14	19				
20					

❖ $cantor_{n+1}(\ \boxed{a_1\ ,\ \dots\ ,\ a_{n-1}},\boxed{an\ ,\ a_{n+1}}\ ) =$

$cantor_n(\ \boxed{a_1\ ,\ \dots\ ,\ a_{n-1}},\ cantor_2(\boxed{an\ ,\ a_{n+1}})\ )$

凡物皆数

- 长度$\boxed{有限}$的字符串，都可视作$\boxed{d = 1 + |\Sigma|}$进制的自然数

  $\boxed{\texttt{decade}} = 453145_{(10)}$

  //d = 1 + ($\boxed{i}$ - $\boxed{a}$) = 10

- 长度$\boxed{无限}$的字符串，都可视作$\boxed{[0, 1)}$内的$\boxed{d}$进制小数

  $\boxed{\texttt{bgahbhahbhdei...}} = 0.2718281828459...$

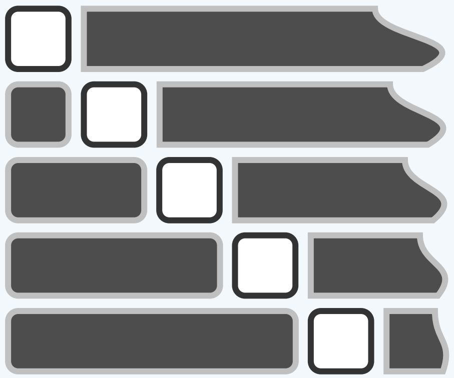
  $\boxed{\texttt{fad}\boxed{\texttt{e}}} = \boxed{0.614\boxed{5}} = \boxed{0.614\boxed{4999...}} = \boxed{\texttt{fad}\boxed{\texttt{diii...}}}$

- $\boxed{\text{Cantor}}$：集合是否$\boxed{可数}$，与其$\boxed{维度}$无关

  有理数集与自然数集一样可数 //(分子，分母)，$\aleph_0$

  无理数集不可数 //$\boxed{\text{Cantor's diagonal}}$，$\boxed{\aleph_0 \boxed{<} \aleph_1}$

ℵ



串亦为数

❖ 十进制串，可直接视作自然数 //指纹（fingerprint），等效于多项式法

P = $\boxed{82818}$　　　　T = 271$\boxed{82818}$284590452353602874713527

❖ 一般地，随意对字符编号{ 0, 1, 2, ..., d - 1 } //设d = |Σ|

于是，每个字符串都对应于一个$\boxed{d}$进制自然数 //尽管不是单射

$\boxed{\text{CAT}}$ = $2\ 0\ 19_{(26)}$ = $1371_{(10)}$ //Σ = { A, B, C, ..., Z }

$\boxed{\text{ABBA}}$ = $0\ 1\ 1\ 0_{(26)}$ = $702_{(10)}$

❖ P在T中出现 $\boxed{\text{仅当}}$ T中某一子串与P$\boxed{\text{相等}}$ //为什么不是$\boxed{\text{当}}$？

❖ 这，不已经就是一个$\boxed{\text{算法}}$吗？！ //具体如何实现？

❖ 问题似乎解决得很顺利，果真如此简单吗？ //复杂度？