中国人民大学
RENMIN UNIVERSITY OF CHINA

第1讲 绪论

余力

buaayuli@ruc.edu.cn

教学小组

- 任课教师：余力
  - 信息学院副教授
  - 办公室：理工配楼406A
  - 电子邮件：buaayuli@ruc.edu.cn
- 助教：刘运涛、马泽宇
  - 信息学院研究生
  - 理工配楼304B

(2021-9)程序设计-余力

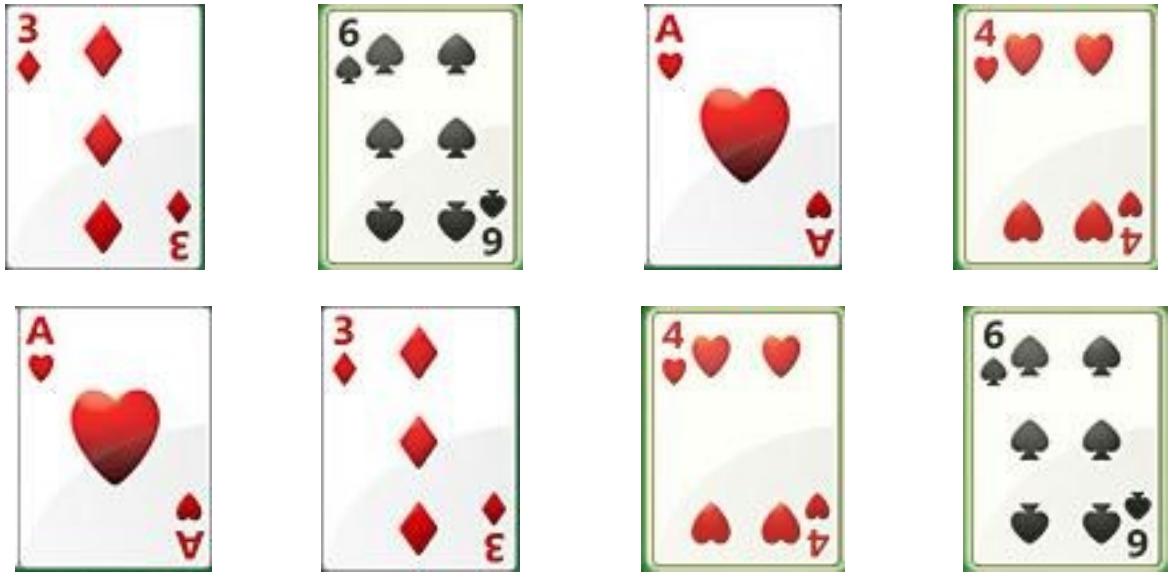
- 上机时间：周二7、8节（下午1、2节）、周五(上午)3、4节
- 上机时间：周五晚 6:00-9:00 明德地下机F90

内容提要

- 1. 计算机怎样理解信息
- 2. 计算机硬件体系结构
- 3. 程序设计简介
- 4. 教学计划与考核要求

案例：扑克排序

- 给你若干张扑克牌，按照牌上数字由小到大的标准排好次序



- 如何让计算机帮你解这一问题？

中国人民大学
RENMIN UNIVERSITY OF CHINA

01. 计算机理解信息

信息与数据

- 从下面例子中比较“信息”与“数据”

在某人的病历卡中：

39℃ ⇒ 数据

体温39℃ ⇒ 信息

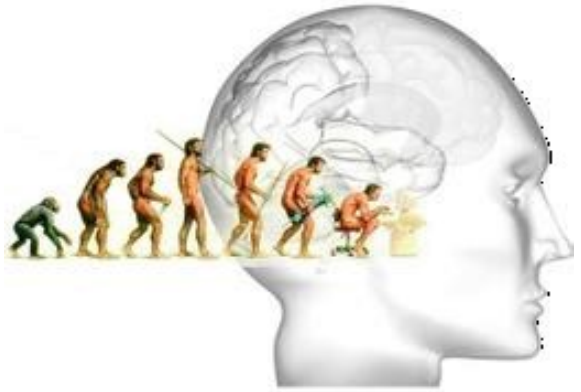
信息　各种事物的变化和特征的反映

数据　信息的载体，如数值、文字、图像等

人理解的是信息，计算机理解的是数据

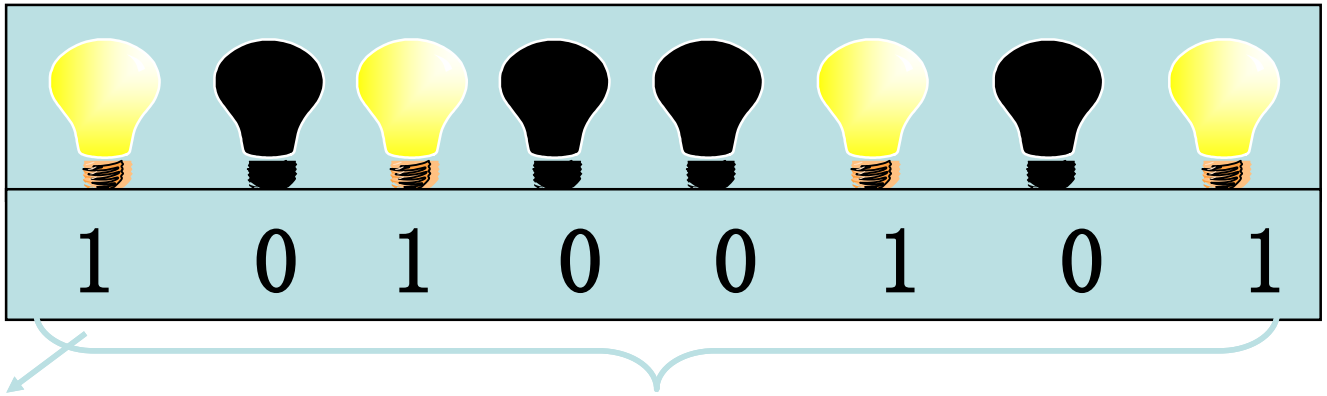
数制：数据的表示方法 (1)

- 数制也称计数制，是指用一组固定的符号和统一的规则来表示数值的方法。
- 几种常用进位数制：
  - 十位制 (Decimal)
  - 二进制 (Binary)
  - 八进制 (Octal)
  - 十六进制数 (Hexdecimal)



二进制：数据在计算机中的表示

- 物理上容易实现，运算简单，可靠性、通用性强
  - 自然界中具有两个固定状态的物理量很多，
  - 例如：电流的有无、电压的高低等。



思考：数值转换

十进制	二进制	十六进制
0	00000000	00
55		
136		
243		
	01010010	
	10101100	
	11100111	
		A7
		3E
		BC

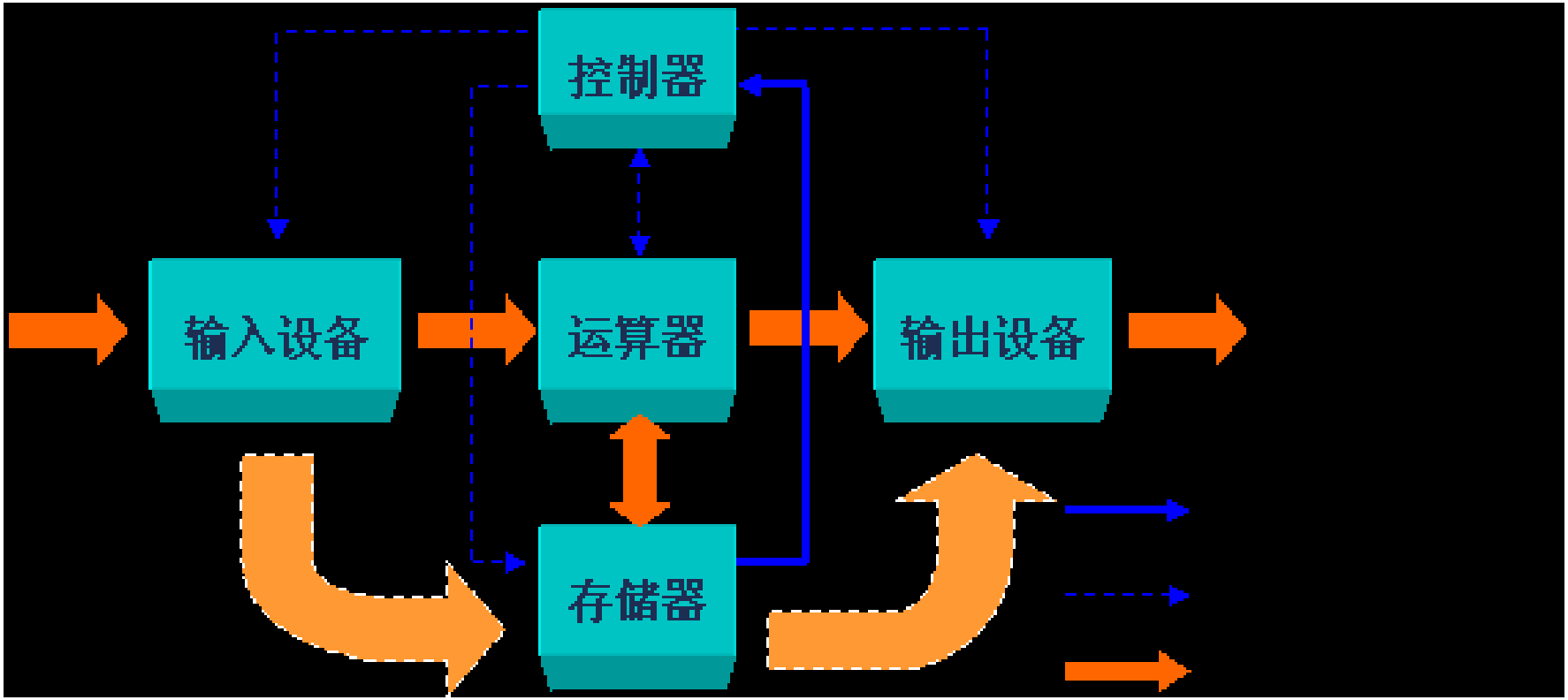
思路：数值转换

- 十进制 → 二进制
  - 除2取余方法
- 二进制 → 十进制
  - 01010010 = 0 * 2^7 + 1 * 2^6 + 0 * 2^5 + ...
- 十六进制 → 二进制
  - 十六进制的1位可以用二进制的几位表示?
- 二进制 → 十六进制
  - 二进制的几位对应着十六进制的1位?

中国人民大学
RENMIN UNIVERSITY OF CHINA

02. 计算机体系结构

冯诺依曼体系结构

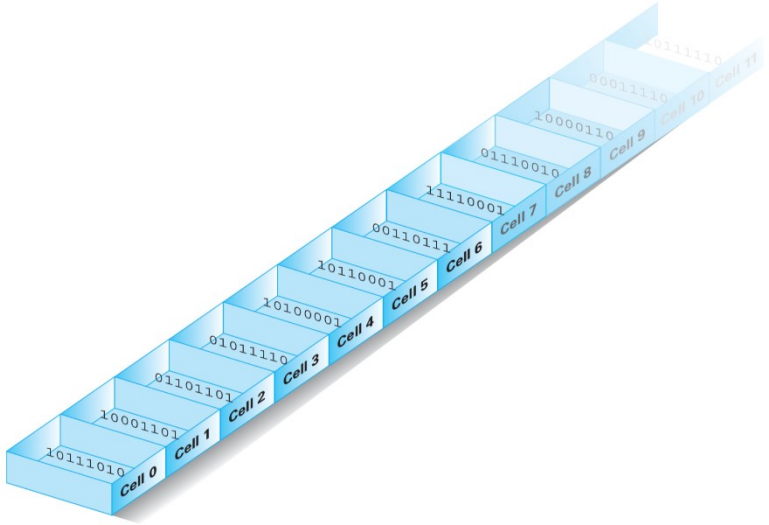


冯诺依曼体系结构 (3)

- 冯诺依曼体系结构的特点
  - 计算机硬件：运算器、控制器、存储器、输入设备和输出设备；
  - 数据和程序以二进制代码形式存放；
  - 控制器根据存放在存储器中的程序来工作
- 控制器
  - 负责从存储器中读取计算机指令，然后解析指令，并按照指令的要求，指挥计算机的其他部件协调工作。
- 运算器
  - 负责执行计算机中的各种运算，从最基本的逻辑运算（与、或、非等），到各种算术运算（加、减、乘、除等）

冯诺依曼体系结构

- 存储器
  - 以二进制形式存储程序和数据。
  - 思考：二进制如何表示整数、小数、字符串、图片、视频等丰富的信息?
- 输入设备
  - 输入设备负责将信息输入计算机
  - 如 键盘、鼠标、扫描仪等;
- 输出设备
  - 输出设备负责将计算结果或存储器中的内容从计算机输出
  - 如显示器、打印机等。



中国人民大学
RENMIN UNIVERSITY OF CHINA

03. 程序设计简介

程序是什么?

Introduction

Social media platforms have attracted billions of users and have emerged as one of the most important channels for companies to communicate with existing and potential customers.[2] Companies have substantially increased their investments and activities in social media marketing,[3] in which visual content plays an important role. The old saying "a picture is worth a thousand words" has never been more true in the era of social media. Images increase the odds of a post getting noticed, especially when people are overwhelmed by the unprecedented amounts of information produced everyday.[4] Industry reports have found concurring results showing that social media posts with images tend to get more likes and shares.[5] Thus a pressing issue for companies conducting social media marketing is to systematically understand the role of visual content in improving customer engagement.[6]

Although social media analysis research has extensively studied textual content (Lakkaraju and Ajmera 2011; Lee et al. 2018; Ma et al. 2015; Stieglitz and Dang-Xuan 2013; Singh et al. 2014) or other characteristics (Lakkaraju and Ajmera 2011; Zadeh and Sharda 2014), visual content has not been systematically investigated due to methodological challenges. In the information systems (IS) or marketing literature, the generation of visual features required significant domain knowledge and manual coding, which is impractical in processing large datasets such as those in social media. In the context of traditional media, studies have shown the impact of images on the effectiveness of ads (Edell and Staelin 1983; Kim and Lennon 2008; Mitchell 1986; Pieters et al. 2010; Pieters et al. 2007; Smith 1991; Tuch et al. 2009; Wu et al. 2016). However, due to methodological restrictions, these studies were based on a small number of hand-picked pictures. Therefore, the findings of these papers can lack generalizability and face scalability issues.

Recently, deep learning emerged as a dominant approach to image recognition in the computer vision field. For example, in the ImageNet image recognition challenge (Russakovsky et al. 2015), performance has dramatically improved due to the breakthrough of neural network-based deep learning approaches. It is believed that, in certain computer vision tasks, deep learning models even surpass human abilities (He et al. 2015; Lake et al. 2015; Rajpurkar et al. 2018; Yu et al. 2016). With such technological advances, we can now generate in a robust and scalable manner visual content features that were previously handcrafted from a small number of images. To the best of our knowledge, a deep learning-based approach has not been widely used to extract visual data features in IS research.

In this research methods article, we introduce deep learning to the social media analysis literature by providing detailed guidance and showing its effectiveness with two empirical case studies. Specifically, we utilize the convolutional neural network (CNN) (Krizhevsky et al. 2017; LeCun et al. 2015), one of the most successful deep learning approaches for visual data analysis. The CNN model aims to automatically discover intricate structure in high-dimensional image data to understand images' semantic content. Our paper provides comprehensive directions on how to conduct a visual content study using deep learning models by summarizing open-source deep learning libraries and outlining steps to construct visual features from images. Additionally, we provide a framework that researchers can use to validate deep learning-induced measures against human coders from Amazon Mechanical Turk (AMT) to ensure their correctness.

In particular, we have generated and validated a group of visual content measures that have important implications in social media research. We first draw on theory from IS and marketing to motivate visual and textual features that can affect consumer information processing, which we operationalize using machine learning (ML) approaches. These features include (1) an object-level image complexity measure based on object detection, (2) an aesthetic score that measures perceived image quality, (3) an adult content score (e.g., underwear, lingerie, etc.), and (4) celebrity endorsements (e.g., Barack Obama, Taylor Swift). In addition, we extract data-driven *generic representations* of visual and textual content that cannot be manually coded (Bengio et al. 2013; LeCun et al. 2015). We then propose novel content features capturing semantic relations between different posts and content types, namely, measures of *content consistency* (e.g., how similar a post is compared to its blog's the average content) and *image-text similarity* (e.g., how closely the image is related to the text within a post). Such measures are difficult to construct without the aid of deep learning approaches, especially for image-text similarity, since pixels and characters are intrinsically different data types.

[2] Statista, "Number of Global Social Media Users" (https://goo.gl/Vxe5uy).

[3] Statista, "Social Media Marketing Spending in the United States from 2014 to 2019" (https://goo.gl/fjQjLeb).

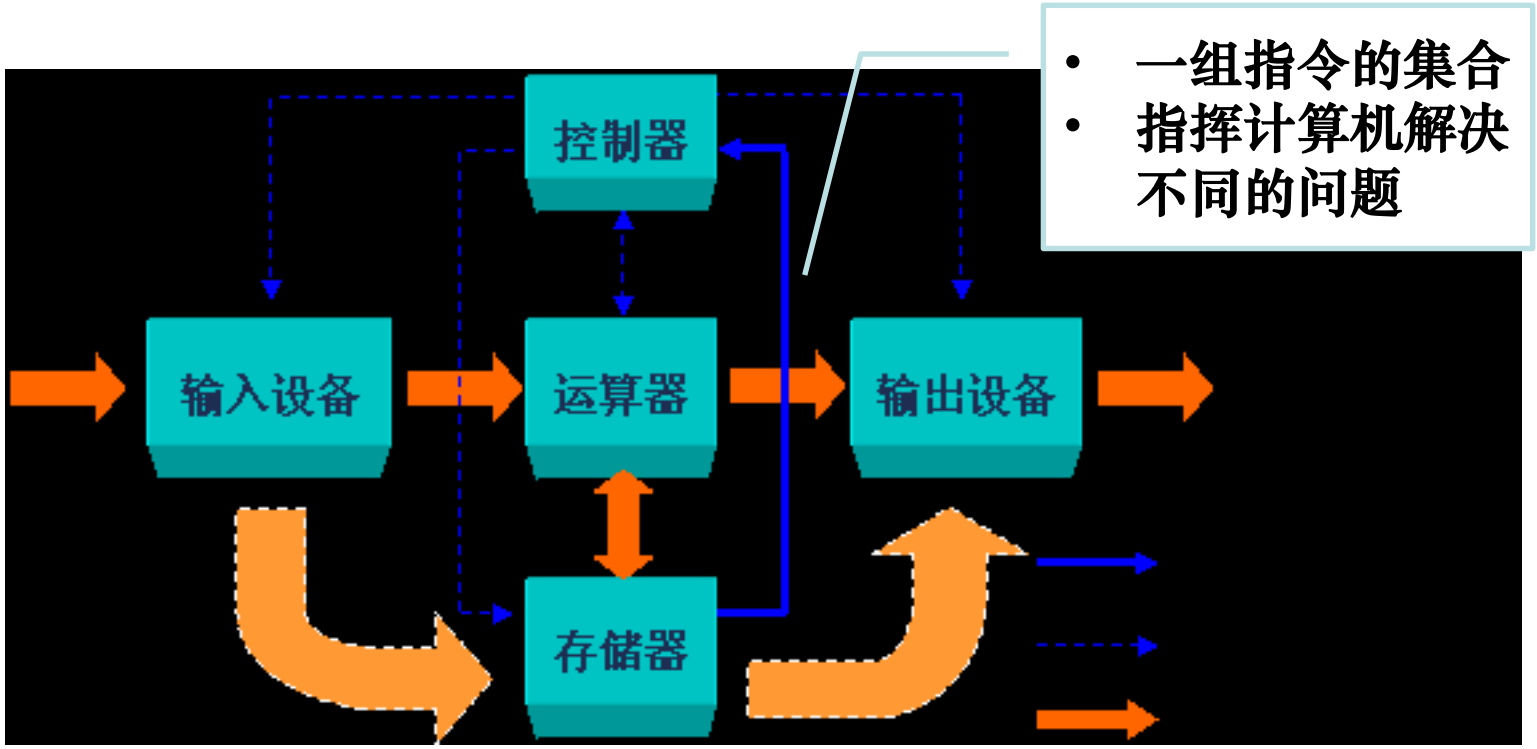
[4] In 2017, Facebook and Instagram users posted 195 million and 95 million new photos daily, respectively (see https://goo.gl/VJY51F).

[5] Adobe Social Intelligence Report, Q1 2014.

[6] HubSpot, "42 Visual Content Marketing Statistics" (https://goo.gl/aRJXs3).

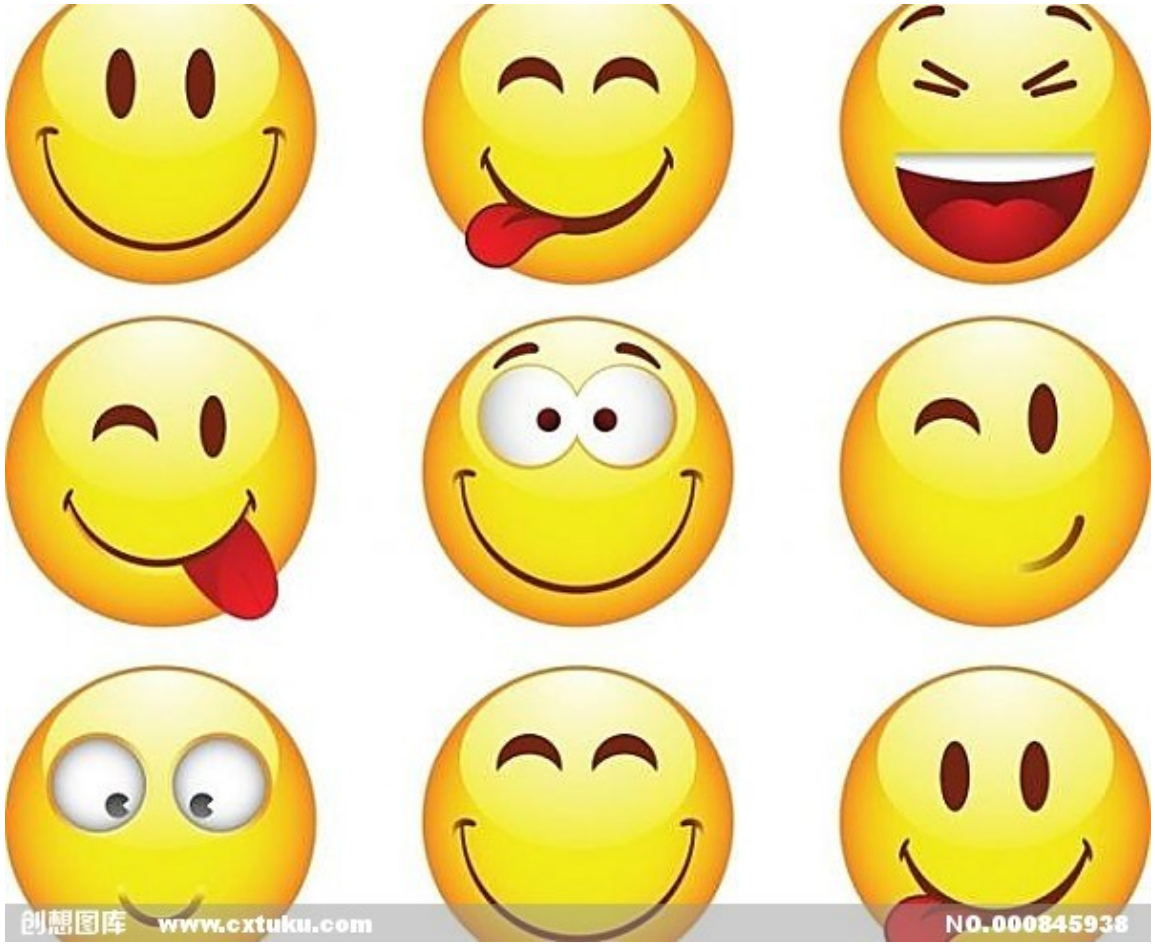
程序是什么?

- 回顾冯诺依曼体系结构



编程：将人的解题步骤变成一组指令

语言是什么?



程序设计语言的发展 (1)

- 机器语言
  - 也称处理器的指令系统，是该处理器可以识别的一组由0和1序列构成的指令码。
  - 指令的格式

操作码OP	地址码AD

  - 指令系统是计算机软件和硬件的界面。
  - 下面是某处理器指令系统中的两条指令：
    - 10000000　加
    - 10010000　减

程序设计语言的发展 (2)

- 汇编语言
  - 50年代中期，人们用一些助记符号来代替0、1码编程。
  - 用助记符号描述的指令系统，称为符号语言或汇编语言。
    - 10000000　加　A+B → ADD A, B

      10010000　减　A-B → SUB A, B
  - 思考：机器能理解如ADD/SUB助记符号?
    - 源(Source)程序 → 目标(Object)程序

程序设计语言的发展 (3)

- 面向过程语言（高级语言）
  - 例子：下面这段C语言程序在做什么？

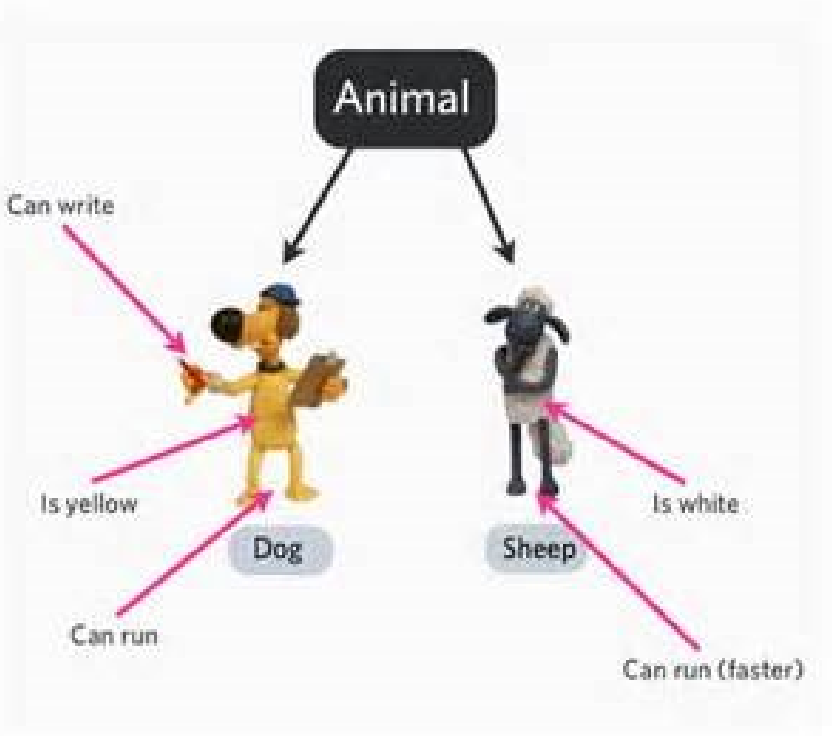
```
#include <stdio.h>

int main() {
        printf("hello, world!\n");
}
```

特点：表达性、紧凑性

程序设计语言的发展 (4)

- 面向对象语言，如Java、C++、Python等



- 新的编程语言不断涌现

中国人民大学
RENMIN UNIVERSITY OF CHINA

04. 教学要求与考核

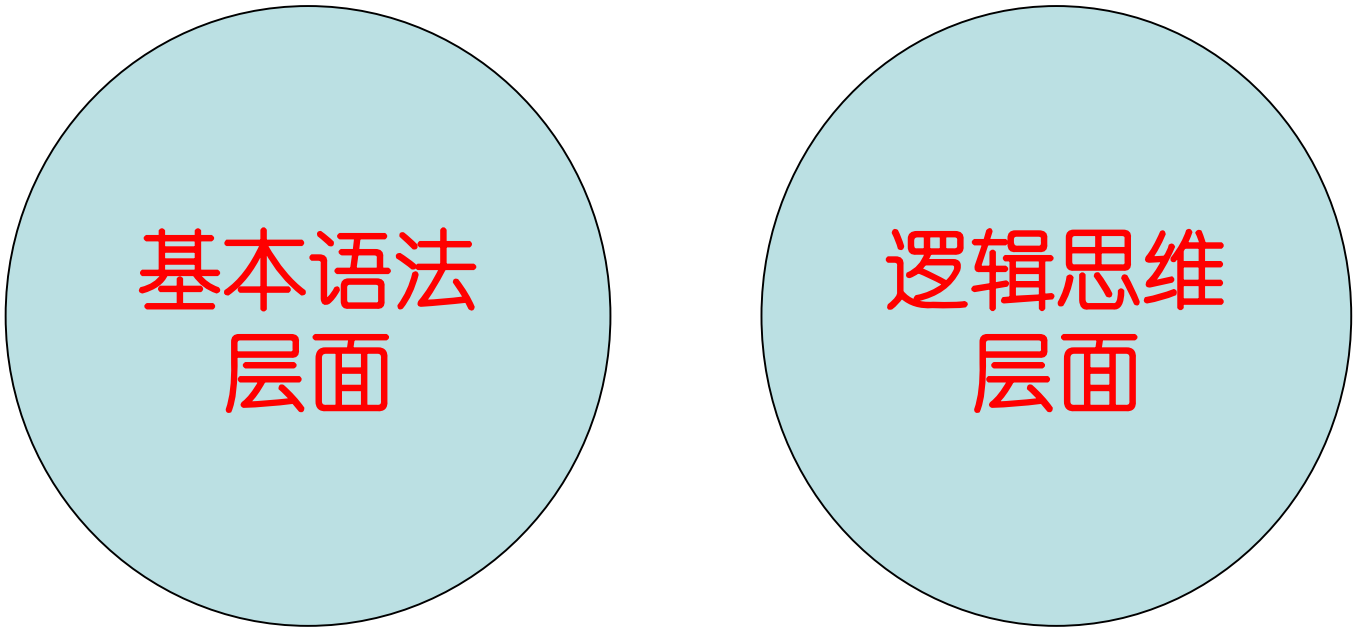
思考

程序设计

==

编程语言学习?

课程目标



学习内容

- 第1讲 绪论及课程介绍
- 第2讲 认识C语言与编程准备
- 第3讲 变量运算与输入输出
- 第4讲 基本程序控制语句
- 第5讲 数组与排序

第6讲 结构体

第7讲 函数

第8讲 递归

第9讲 指针

第10讲 文件

程序设计课怎么学?

你认为的程序设计

实际上的程序设计

程序设计是练会的

程序设计课怎么学?

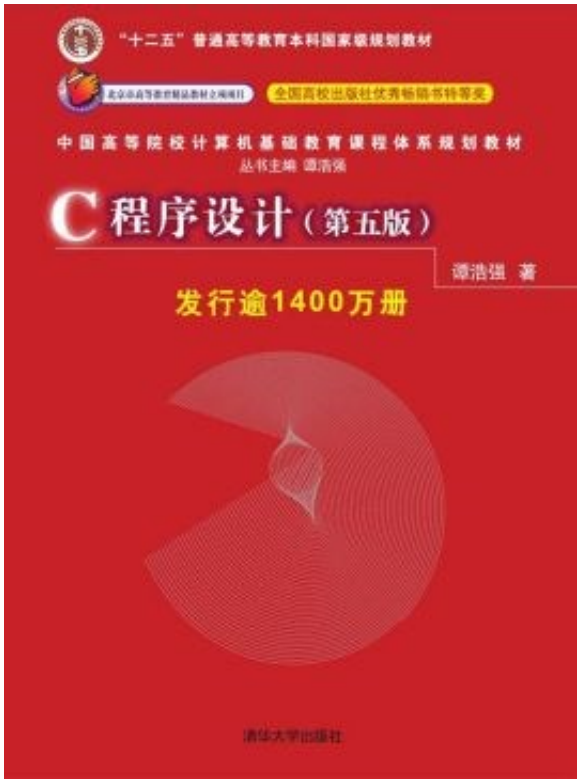
- 上机是核心
  - 上机5道题（ Dealine是当天，**周六**助教检查上机）
  - 课后3道题（ Dealine是周二，**周三**助教检查作业）
  - 一学期下来约100道
  - 上机要及时记录问题（当天交）
- 课程学习互助小组
  - 5人一组，固定上机坐位在一起
  - 交流讨论问题
  - 组长负责汇总小组上机问题情况，交给助教
- 出现问题怎么办
  - 个人 - 小组 - 助教 - 老师

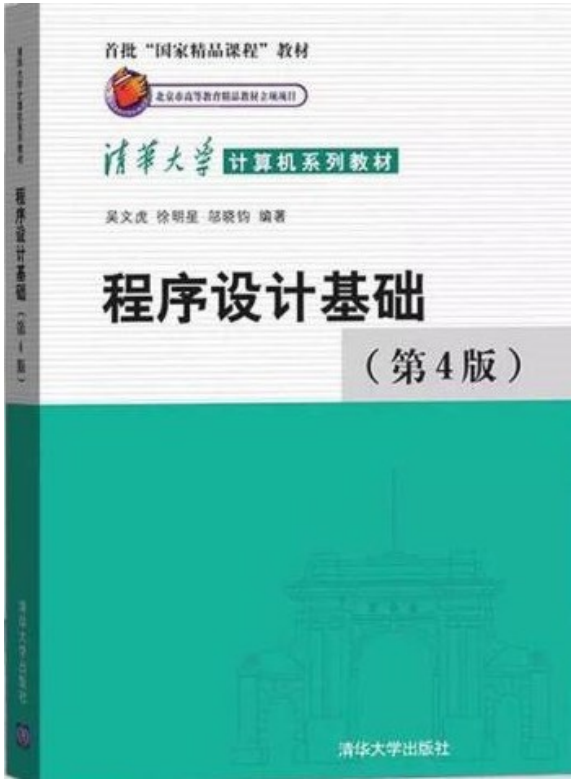
程序设计导论课怎么考？

- 期末考试：50%
  - 统一上机考试：计算机评分
- 平时成绩：50%
  - 期中考试：20%（计算机评分）
  - 平时作业：20%（友学系统）
  - 课堂表现：10%

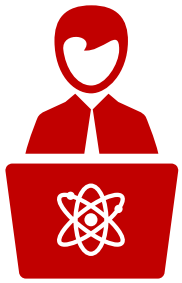
参考教材

- **谭浩强．C程序设计（第五版）. 清华大学出版社.**
- 吴文虎 等. 程序设计基础（第四版）. 清华大学出版社.





中国人民大学
RENMIN UNIVERSITY OF CHINA



谢谢大家！