

# 信息表达与智能处理

## （人工智能知识表达基础）

2019年招生专业“电子信息科学与技术 (AI+)”

电子工程学院 通信与网络实验室

范春晓

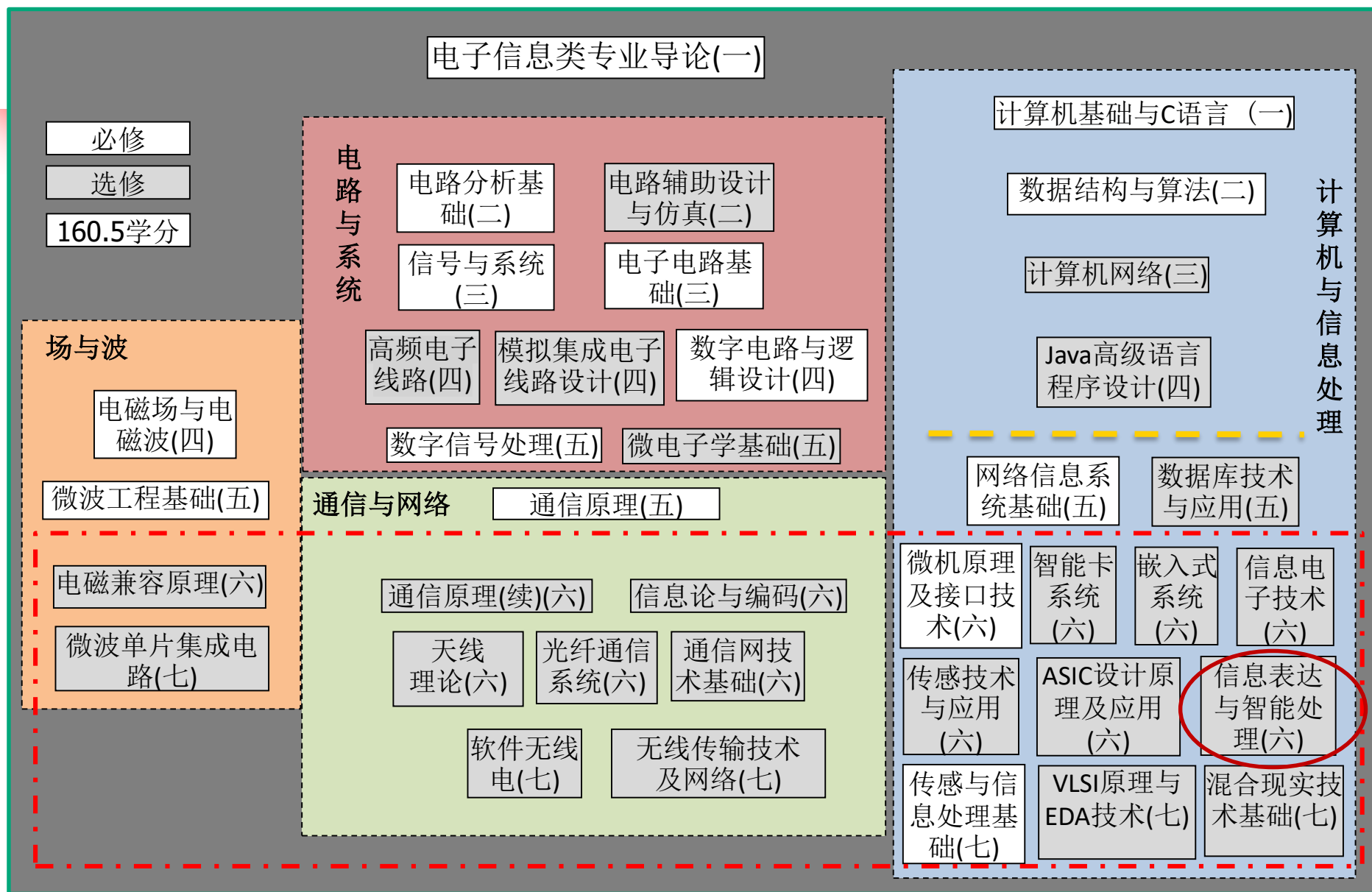
[fcxg100@163.com](mailto:fcxg100@163.com)

教师办公室：教四242 Tel: 62283467

实验室主页：[www.buptcnc.cn](http://www.buptcnc.cn)

2021-3

# 电子信息科学与技术专业关系图（2017-）





# 信息表达与智能处理

---

## ■ 主要参考资料

- 课件
- 《人工智能导论(第3版)》 丁世飞编著 中国工信出版集团 电子工业出版社 2020年1月
- 《走进搜索引擎》 潘雪峰 花贵春 梁斌编著 电子工业出版社 2011年5月第2版
- 《Python带我起飞 入门、进阶、商业实战》 李金洪编著 中国工信出版集团 电子工业出版社 2019年9月第3次印刷
- 《Python 爬虫开发与项目实战》 范传辉编著 机械工业出版社 2017年11月第一版



# 信息表达与智能处理

---

- 先修课：网络信息系统基础、数据结构
- 2021春季开课说明
  - 网上开课方案：腾讯课堂
  - 备选方案：微信群，PPT录播
  - 作业及答疑：微信群，邮箱：fcxg100@163.com
- 考核方式
  - 考查（开卷、随堂）/（大作业）



# 信息表达与智能处理

---

第1章 概述

第2章 万维网网页信息的表达及解析

第3章 Python语言入门与网络爬虫

第4章 知识表示方法

第5章 知识图谱及资源描述框架RDF

第6章 正则表达式与Json、ASN.1

第7章 语义Web本体语言OWL



# 第1章 概述

---

1.1 数据、信息、知识

1.2 数据表示与数据结构

1.3 信息表示规范及层次

1.4 知识表示与信息智能处理





# 1.1 数据、信息、知识

- 数据：事务、概念或指令的一种形式化的表示形式，以便于人工或自然方式进行通信、解释或处理。
  - 连续的值，比如声音、图像，称为模拟数据
  - 离散的，如符号、文字，称为数字数据。
- 信息：数据所表达的客观事实。
  - 数据是信息的载体，与具体的介质和编码方法有关
- 知识：
  - 经过加工的信息，包括事实、信念和规则
  - 人们把实践中获得的信息关联在一起，就形成了知识
  - 把有关信息关联在一起所形成的信息结构





# 1.1 数据、信息、知识

---

- 知识的特点：

- 人通过实践，认识到的客观世界的规律性的东西
- 在信息的基础上增加了上下文信息，提供了更多的意义，更加有用和有价值

- 例.在人工智能中的知识：“规则”，反映了信息之间的因果关系

- 用得最多的关联是“如果... ..，则... ..”
- 如果大雁向南飞，则冬天就要来了

- 例.在人工智能中的知识：“事实”

- 雪是白色的



# 1.1 数据、信息、知识

- 数据、  
信息、  
知识例

- 数据：37.5

- 信息：

- 测量体温为37.5度

- 知识：

- 是不是发烧，要不要去医院？！

- 数据：

- 通过观察可以记录星球的位置和出现的时间等原始数据

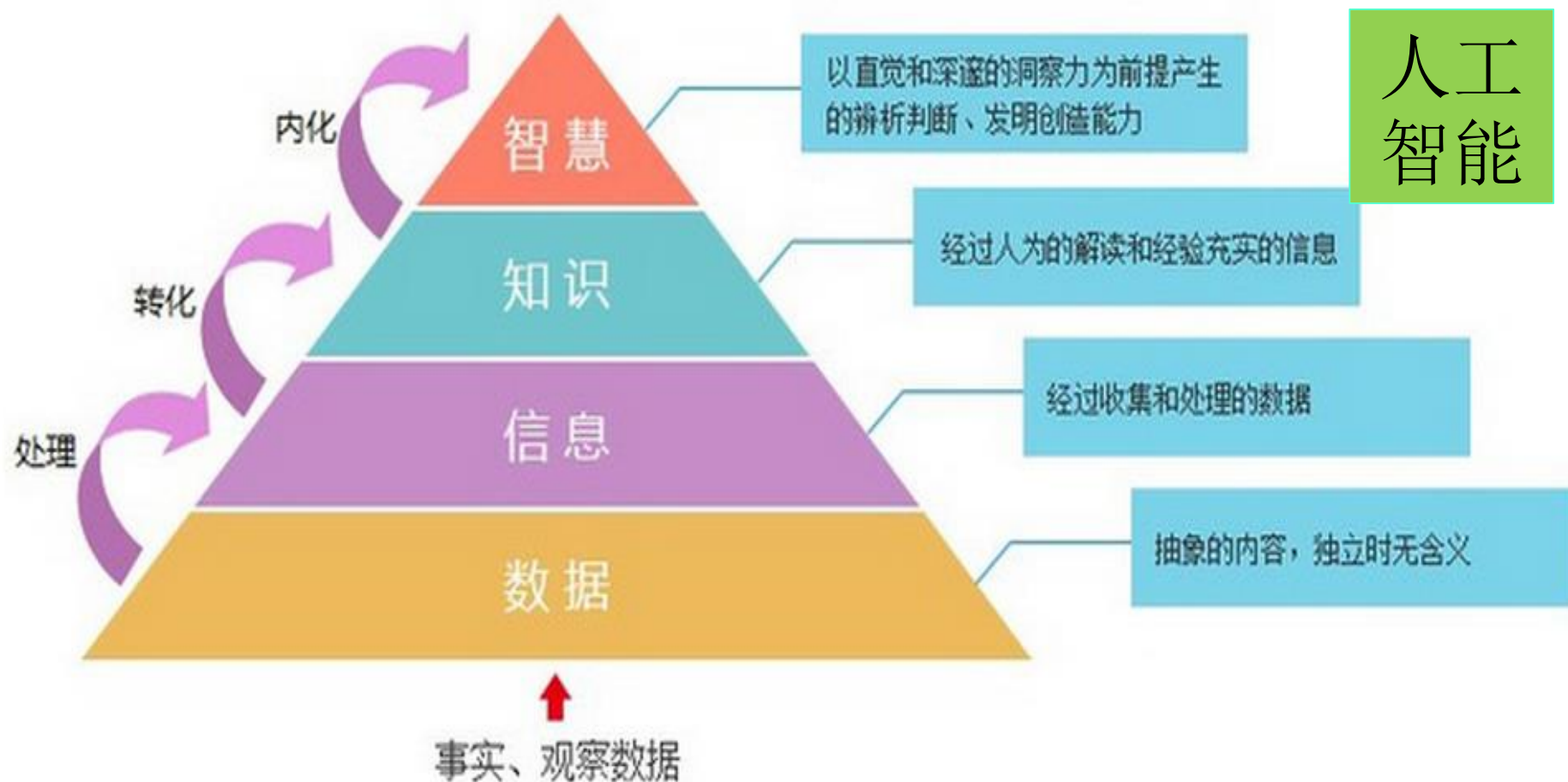
- 信息：

- 对得到的数据进行分析 and 挖掘，就能够计算出星球运动的规律

- 知识：

- 对信息进行总结和提炼，就可以得到开普勒三定律

# 1.1 数据、信息、知识





# 第1章 概述

---

1.1 数据、信息、知识

1.2 数据表示与数据结构

1.3 信息表示规范及层次

1.4 知识表示与信息智能处理

## 1.2 数据表示与数据结构

### ■ 数据表示：

- 为便于计算机存储和使用，国际标准组织规定了各类数据符号的表示标准



## 1.2 数据表示与数据结构

- 数据组织 → 存储 → 共享 → 需要标准（例）

### (1) 数值型数据

- 二进制编码表示
- 用0和1的序列表示数值及符号



二进制浮点数算术标准（ANSI/IEEE Std 754-1985）  
又称IEC 60559:1989，微处理器系统的二进制浮点数  
算术



# 1.2 数据表示与数据结构

## ■ (2).字符数据的编码表示

- 美国信息交换标准代码ASCII（American Standard Code for Information Interchange）ISO/IEC 646
  - 一种标准的单字节字符编码方案，用于基于文本的数据，使用指定的7 位或8 位二进制数组合来表示128 或256 种可能的字符
- Unicode（统一码、万国码、单一码）
  - 为了解决传统的字符编码方案的局限而产生的
  - 为每种语言中的每个字符设定了统一并且唯一的二进制编码，以满足跨语言、跨平台进行文本转换、处理的要求。1990年开始研发，1994年正式公布
  - Unicode通常用两个字节表示一个字符，原有的英文编码从单字节变成双字节

### ■ (3).汉字的编码表示

#### ■ 汉字交换码：

- 连续两个字节（16个二进制为）表示一个汉字，GB2312-80，6763个汉字，682个符号，共7445个字符
- GB2312未考虑与ASCII冲突，所以不能直接在计算机内使用

#### ■ 汉字机内码：

- 计算机内表示汉字
- 把交换码两个字节最高位改为1，成为机内码

#### ■ 汉字字形码：用来将汉字显示到屏幕或打印纸上所需的图形数据

- 点阵法（点阵码）：把汉字按字形排列成点阵，一个16X16点阵占用32个字节
- 矢量法（矢量码）

#### ■ 汉字输入码：将汉字通过键盘输入到计算机中采用的代码，也称为汉字外部码（外码）：

- 拼音码、拼形码、音形码、序号码



# 1.2 数据表示与数据

## (4).图形和视频表示

### ■ 位图

- 计算机屏幕是由一个个像素组成的，的黑白不同则组成一幅幅计算机黑白画
- 存储黑白位图，用1代表白，0代表黑
- 分辨率800X600的屏幕有480000个像素，需要480000个比特存储一个黑白位图

### ■ 矢量图

- 由一系列可重构图像的指令组成
- 在创建矢量图时，计算机把不同颜色的线和图形转换成能够重构的指令，计算机存储的是指令而非真正的图像
- 优点：存储空间比位图小，不随图像大小而变

点阵图



放大后的点阵图



矢量图



放大后的矢量图



思缘教程网

www.missy.cn



# 1.2 数据表示与数据结构

## (4).图形和视频表示（续）

- 位图灰度图像
  - 用不同程度的灰色表示位图，灰度的级别越多，图像越真实
  - 计算机常用256级灰度显示图像，每个像素可以选择256级灰度中的一种
  - 存储量增加，256级灰度需8比特，黑白图像仅需1比特
- 位图彩色图像
  - 有16色、256色、24位图像（真彩色图像），分别可选择16种、256种和 $2^{24}$ 种色彩
  - 每个像素分别需要4个比特、8个比特和24个比特

## 1.2 数据表示



### ■ (4).图形和视频表示 (续)

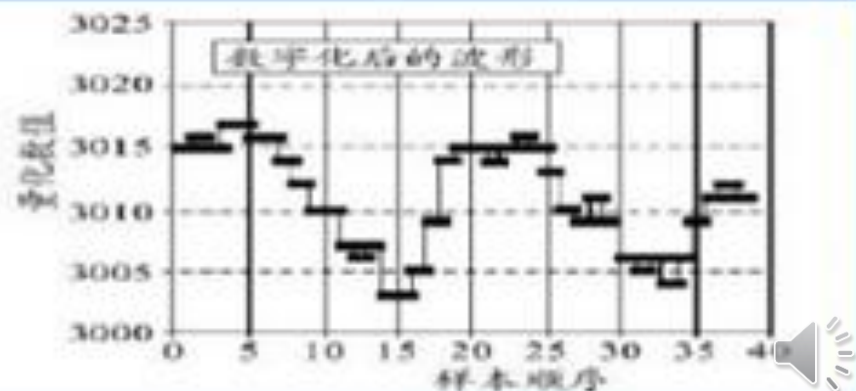
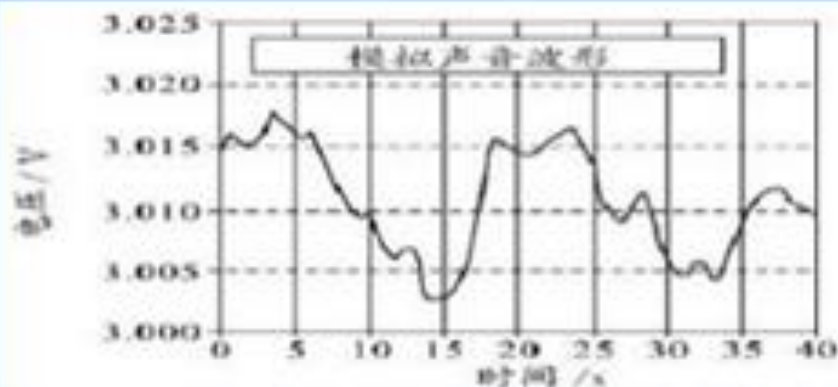
#### ■ 数字视频

- 由一系列的帧组成，每一帧都是一幅静止的位图图片
- 视频每秒钟显示**30**帧图片
- 位图是静态的，视频是动态的，这是由于人眼看到每一幅图像时会产生视觉暂留
- 如果在视觉暂留褪去之前显示下一幅图片，人们就会感觉到动态的效果。

# 1.2 数据表示与数据结构

## (5).声音表示

- 模拟方式和数字方式
- 音频信号：模拟量，当声音传递到话筒时，话筒将声音引起的震动转换为电信号，用不同的电压表示声音大小，即用电压波动来模拟声波，这种表示声音的电信号称为模拟音频。
- 模/数转换：模拟量 $\rightarrow$ 数字（数字音频），便于计算

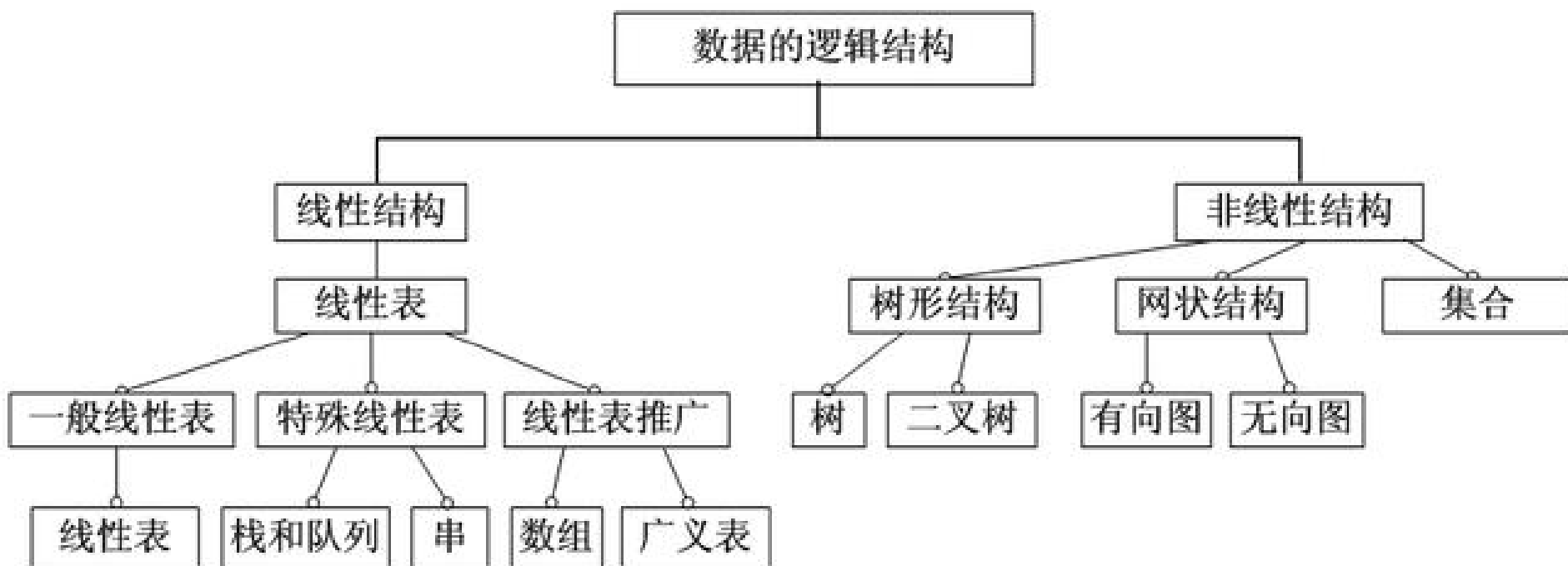




# 1.2 数据表示与数据结构

## 2. 数据结构

- 数据结构是计算机存储、组织数据的方式。
- 数据结构是指相互之间存在一种或多种特定关系的数据元素的集合。
  - 包括三个组成成分
    - 数据的逻辑结构:指数据元素之间的前后间关系, 而与他们在计算机中的存储位置无关
    - 数据的存储结构:数据的逻辑结构在计算机存储空间的存放形式(顺序、链接、索引、散列等)
    - 数据运算结构
- 精心选择的数据结构可以带来更高的运行或者存储效率



# 数据结构应用场景（特点）

数据结构	优点	缺点
数组	插入快	查找慢、删除慢、大小固定
有序数组	查找快	插入慢、删除慢、大小固定
栈	后进先出	存取其他项很慢
队列	先进先出	存取其他项很慢
链表	插入、删除快	查找慢
二叉树	查找、插入、删除快	算法复杂（删除算法）
红黑树	查找、插入、删除快	算法复杂
hash表	存取极快（已知关键字）、插入快	删除慢、不知关键字时存取很慢、对存储空间使用不充分
堆	插入快、删除快、对大数据项存取快	对其他数据项存取慢
图	依据现实世界建模	算法有些复杂





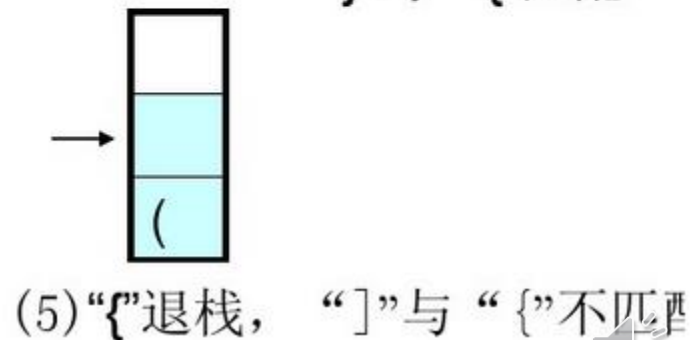
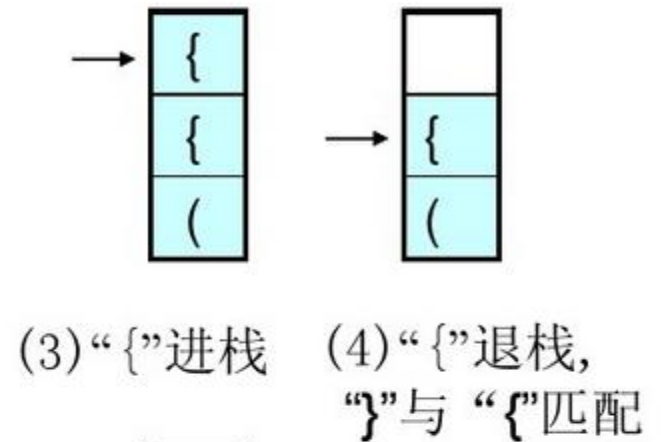
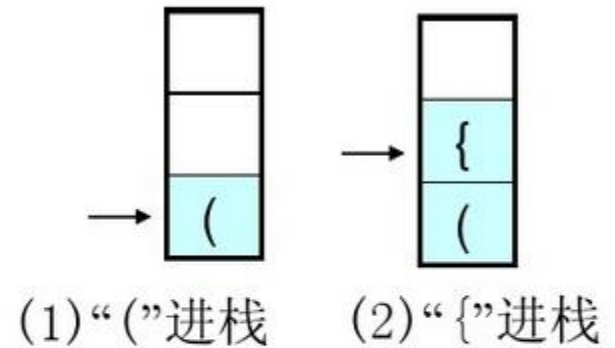
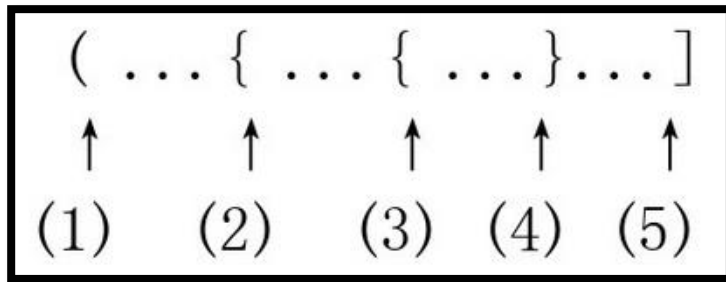
## 1.2 数据表示与数据结构

- 例.括号配对，数据结构的经典题目，给你一串由括号组成的字符串，问其中所有括号是否满足两两配对
- 思路：

先选数据结构：栈

  - 遍历字符串
  - 如遇左括号，则入栈
  - 遇到右括号，则比较该右括号是否与栈顶配对
    - 是，即出栈，继续查找下一个字符
    - 否，括号不匹配，提前结束遍历
  - 最后结束遍历以后，若栈为空，则括号配对，否则不配对





```

Status matching(string& exp) {
    int state = 1;
    while (i <= Length(exp) && state) {
        switch of exp[i] {
            case “(”:{Push(S,exp[i]); i++; break;}
            case “)”:{
                if(NOT StackEmpty(S)&&GetTop(S)=“(”
                    {Pop(S,e); i++;}
                else {state = 0;}
                break; } ... ..
        }
    }
    if (StackEmpty(S)&&state) return OK; .....
}
  
```

## 1.2 数据表示与数据结构

### 行编辑程序

功能：接受用户从终端输入的程序或数据，并存入用户的数据区。  
退格符“#”：表示前一个字符无效。

#### ■ 思路：

先选数据结构：栈

- 遍历字符串
- 如遇非#字符，则入栈
- 遇到#字符，弹出栈顶字符
- 最后结束遍历以后，栈中即为有效字符串

whli # # ilr # e(s # \*s)

while (\*s)

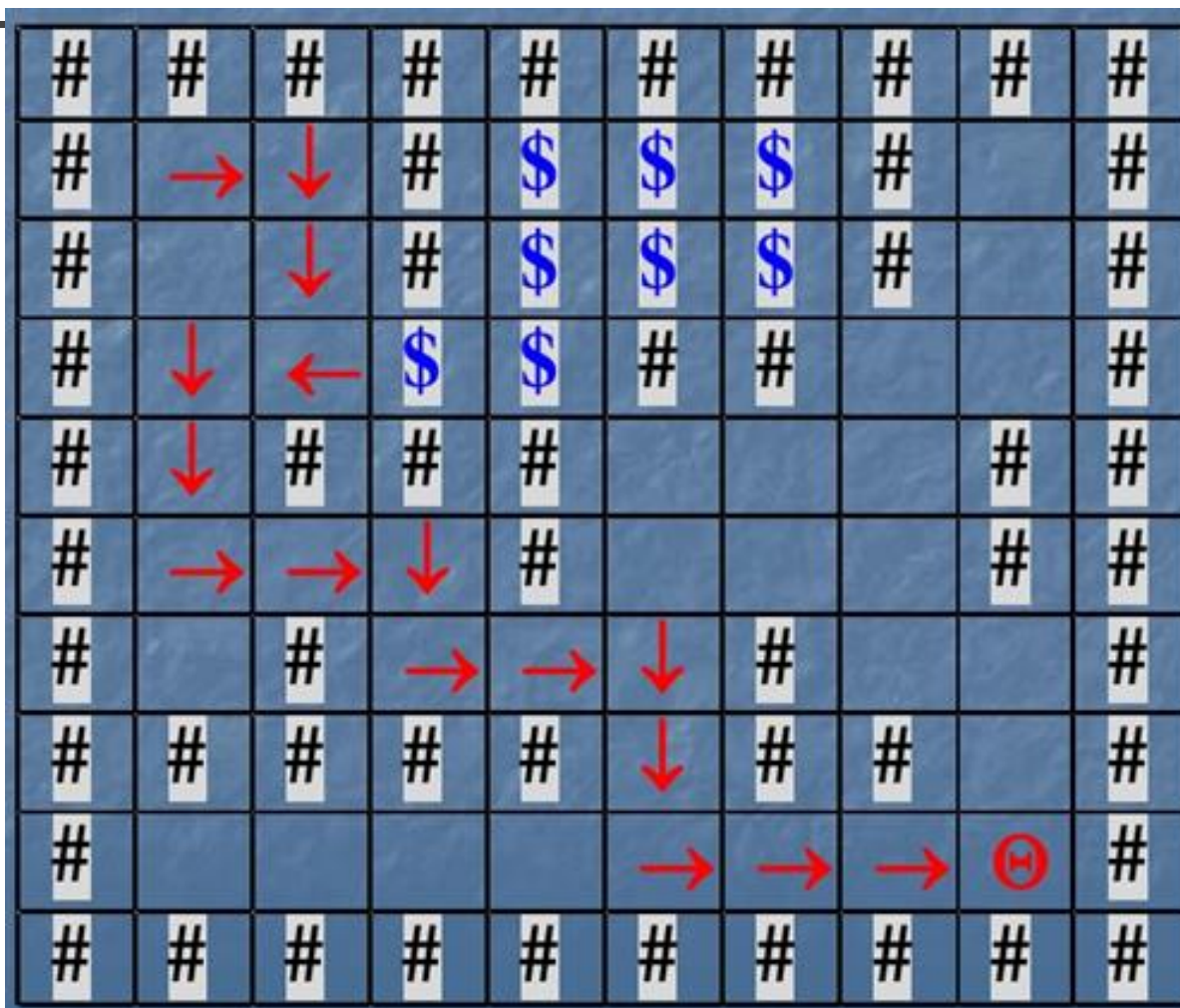
算法3.2如下:

```
void LineEdit { //利用字符栈S, 从终端接收一行并传送至调用过程的数据区。  
    InitStack (S);           //构造空栈S  
    ch = getchar ();         //从终端接收第一个字符  
    while (ch != EOF) {      //EOF为全文结束符  
        while (ch != EOF && ch != '\n') {  
            switch (ch) {  
                case '#': Pop (S, c);          break; //仅当栈非空时退栈  
                case '@': ClearStack (S);      break; //重置S为空栈  
                default: Push (S, ch);         break; //有效字符进栈,  
                                                    //未考虑栈满情况  
            }  
            ch = getchar ();                   //从终端接收下一个字符  
        }  
        将从栈底到栈顶的栈内字符传送至调用过程的数据区;  
        ClearStack (S);                      //重置S为空栈  
        if (ch != EOF)  
            ch = getchar ();  
    }  
    DestroyStack (S);  
} //LineEdit
```



# 数据结构迷宫求解

<https://wenku.baidu.com/view/707bca52876fb84ae45c3b3567ec102de2bddf02.html>





# 迷宫的分析

## ■ 数据结构是数组和栈

- 迷宫设置为一个2维数组，通路为1，不通为0，但是四周为屏障
- 设置一个栈来存储迷宫的路径：记录每个位置的坐标值  $(x,y)$ ，同时将纳入栈中的路径，记录它来自何方，也就是记录它的探测方向编号（东，南，西，北类似于地图的指示，0，1，2，3）

## 迷宫算法

判断  
当前  
节点  
是否  
畅通

### ■ 通的话就入栈操作： $(x,y,di)$

通畅，则记录该点到栈中，并判断是为终点，不为终点的话，继续前行探索

设置访问东邻居开始，若其不通，换方向探路

邻居访问遍，均不通，退出当前结点，取当前栈顶的未访问邻居探路

不畅通，则后退，换方向访问，到栈空结束



# 第1章 概述

---

1.1 数据、信息、知识

1.2 数据表示与数据结构

1.3 信息表示规范及层次

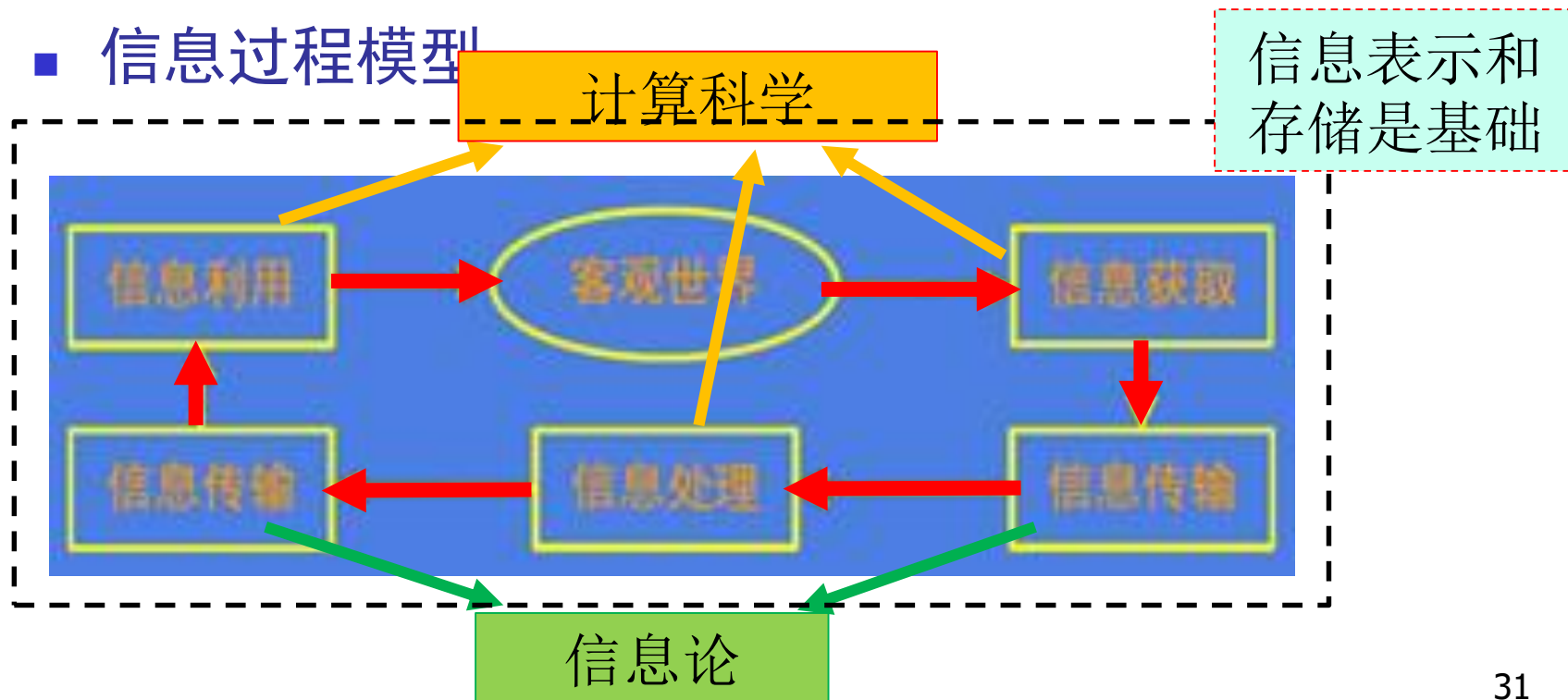
1.4 知识表示与信息智能处理

# 1.3 信息表示规范及层次

## ■ 信息技术:

- 信息获取、信息传输、信息处理、信息存储和信息利用的，以信息为主要研究对象的所有科学与技术

## ■ 信息过程模型





# 1.3 信息表示规范及层次

## ■ 信息数据（计算机科学）

### ■ 信息数据范畴：

- 计算机出现之前：数字、字符
- 互联网出现之前：文字并不被认为是数据
- 互联网出现之后：聊天记录、网页内容、打电话记录、用户生存内容（**UGC**）、论坛评论，购物数据，社会关系，行程记录等等都是数据内容。

非结构化  
结构化  
半结构化







# 1.3 信息表示规范及层次

## ■ 信息层次：

- 感测技术（信息获取）
- 通信技术（信息传输）
- 计算机技术（信息处理）
- 控制技术（信息利用）

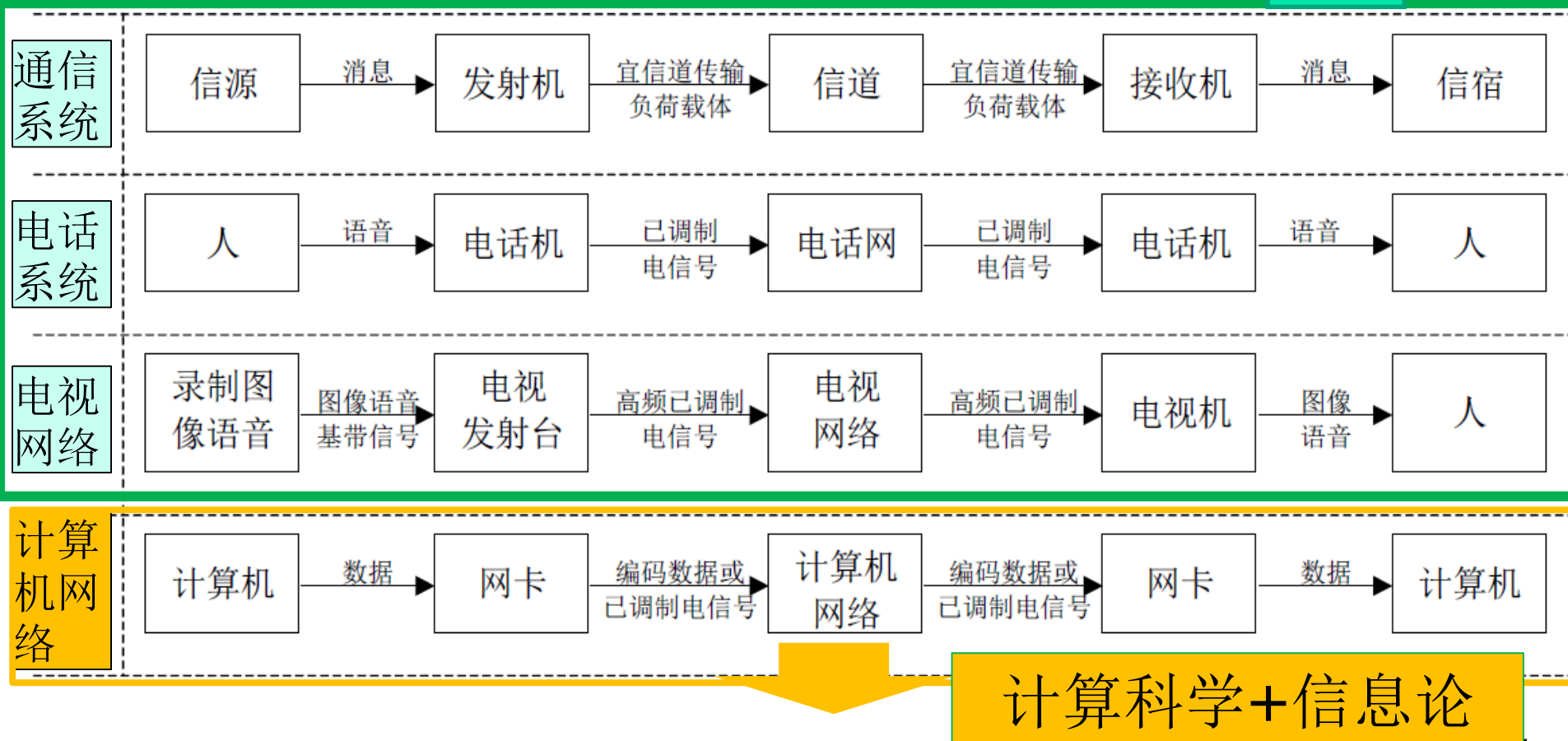
任意层次的信息表示要素：  
内容元素  
信息单元  
信息集合体系

- 不同层次信息作用不同，需要的信息不同；
- 当前信息技术的基本特征就是计算机程序控制化
- 信息表示是基础
  - 目标：开放的、计算机可识别处理
  - 作用：存储、交换和互操作

# 1.3 信息表示规范及层次

信息论

## ■ 信息传输





## 1.3 信息表示规范及层次

### ■ 计算机科学中信息表示层次涉及的规范：

- 信息获取：感知、检测
- 信息处理：

非结构化数据、二进制文件、十六进制文件等

- 存储：适合计算机存储
- 交换：**ASN.1**、**XML Scheme**、**JSON**
- 结构化：数据库、数据结构
- **Web**表达：**HTML**、**XML**、**RDF**、**JSON**、**CSS**
- 检索、搜索：**正则表达式**、排序、倒排、...
- 信息利用：
  - 语义：**本体**、**OWL**
  - 知识表示：知识图谱、知识库



## 1.3 信息表示规范及层次

- 利用计算机进行信息交流时，事先必须对各类信息制定统一的“编码”标准，使得通过计算机以及网络交流信息成为可能。
- 目前国际公认的信息表示规范有：
  - 英文字符信息交换的ASCII码；
  - 汉字信息交换的国标码GB2312；
  - 商品信息的条形码；
  - 网络数字音乐的MP3编码；
  - 静态图像压缩技术的JPEG标准和视频压缩技术MPEG标准。
  - 等等



# 第1章 概述

---

1.1 数据、信息、知识

1.2 数据表示与数据结构

1.3 信息表示规范及层次

1.4 知识表示与信息智能处理



# 1.4 知识表示与信息智能处理

## ■ 知识表示

- 研究用机器表示知识的可行性、有效性的一般方法，可以看成将知识符号化，即编码成某种数据结构，并输入到计算机的过程和方法
- 知识表示就是将人类知识形式化或者模型化
- 目的：让计算机存储和运用人类知识

知识表示=数据结构+处理机制

- 常用表示法：
  - 一阶谓词逻辑表示法
  - 产生式表示法
  - 框架表示法
  - 语义网络表示法
  - 脚本表示法
  - 面向对象表示法



# 1.4 知识表示与信息智能处理

- 智能本意，两个含义：

- 名词：指人类所能进行的脑力劳动，包括感觉、认知、记忆、学习、联想、计算、推理、判断、决策、抽象、概括等
- 形容词：人一样的、聪明的、灵活的、柔性的、自学习的、自组织的、自适应的、自治的等等。

- 智能理论的研究，两个方面：

- 自然智能理论：
  - 对智能的产生、形成和工作的机制直接研究；
  - 生理学和心理学研究者从事；
- 人工智能理论： 信息技术发展方向是智能化
  - 如何用人工的方法模拟、延伸和扩展智能；
  - 理工学研究者从事。



# 1.4 知识表示与信息智能处理

- 人工智能（Artificial Intelligence）四类定义：
  - 类人行为系统
  - 类人思维系统
  - 理性思维系统
  - 理性行为系统
- 人工智能主要研究用人工的方法和技术，模仿和扩展人的智能，实现机器智能。
- 人工智能主要领域：

自然语言处理  
知识表示  
自动推理  
机器学习  
计算机视觉  
机器人技术

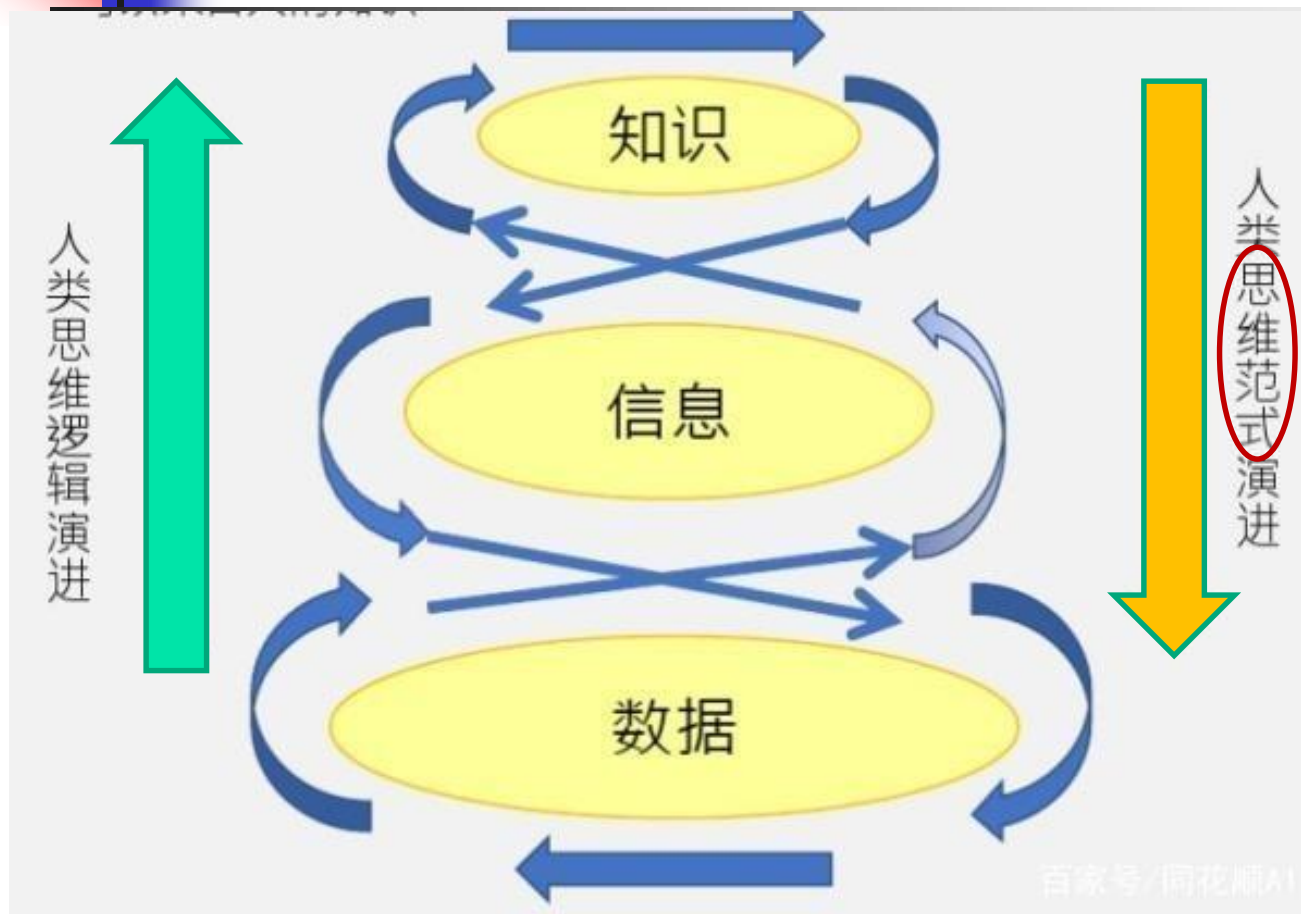
智能系统-  
信息智能处理

知识表示是智能系统的重要基础

- 为了使计算机具有智能行为，就必须使它具有知识
- 知识需要适当的模式表示才能存储到计算机中



## 1.4 知识表示与信息智能处理



### 思维范式

- 是指立足于一种世界观、认知体系、信念等而形成的固有的、稳定的、反复使用的具有范例特点的思维规范、模型或模式。

# 1.4 知识表示与信息智能处理

## ■ 信息智能处理应用场景



# 1.4 知识表示与信息智能处理

## ■ 信息智能处理——机器人



# 1.4 知识表示与信息智能处理

## ■ 信息智能处理——智能手机与非智能手机



# 1.4 知识表示与信息智能处理

## ■ 信息智能处理——汉字输入、语音识别





# 1.4 知识表示与信息智能处理

## 模式识别

利用机器对物体、图像、语音、字符等进行自动识别的技术。

图像识别

文字识别

人脸识别

语音识别(转写、声纹识别等)

## 自然语言处理

语音识别

机器翻译

听懂人的、图片上的等等给定信息的含义

权威词典

海量词句



# 信息智能处理——搜索引擎

Baidu 百度

姚明



百度一下

Q 网页

资讯

视频

图片

知道

文库

贴吧

地图

采购

更多

百度为您找到相关结果约100,000,000个

搜索工具

[姚明\(亚洲篮球联合会主席、中国篮球协会主席\) - 百度百科](#)



生涯: 篮球

生日: 1980年9月12日

专业特点: 20英尺外精确跳投

主要成就: 8次NBA全明星 (2003-2009; 2011), ESPN全球最有潜...

[早年经历](#) [职业生涯](#) [比赛数据](#) [个人生活](#) [主要作品](#) [更多 >](#)

[baike.baidu.com/](http://baike.baidu.com/)

[姚明 - 海量精选高清图片 - 百度图片](#)

[全部](#) [赛场](#) [壁纸](#) [公益](#) [生活照](#) [活动](#) [代言广告](#) [节目](#) [亲密关系](#) [粉丝制作](#) [写真](#)



相关篮球运动员



叶莉

前中国著名篮球运动员



方凤娣

姚明母亲原篮球运动员



王治郅

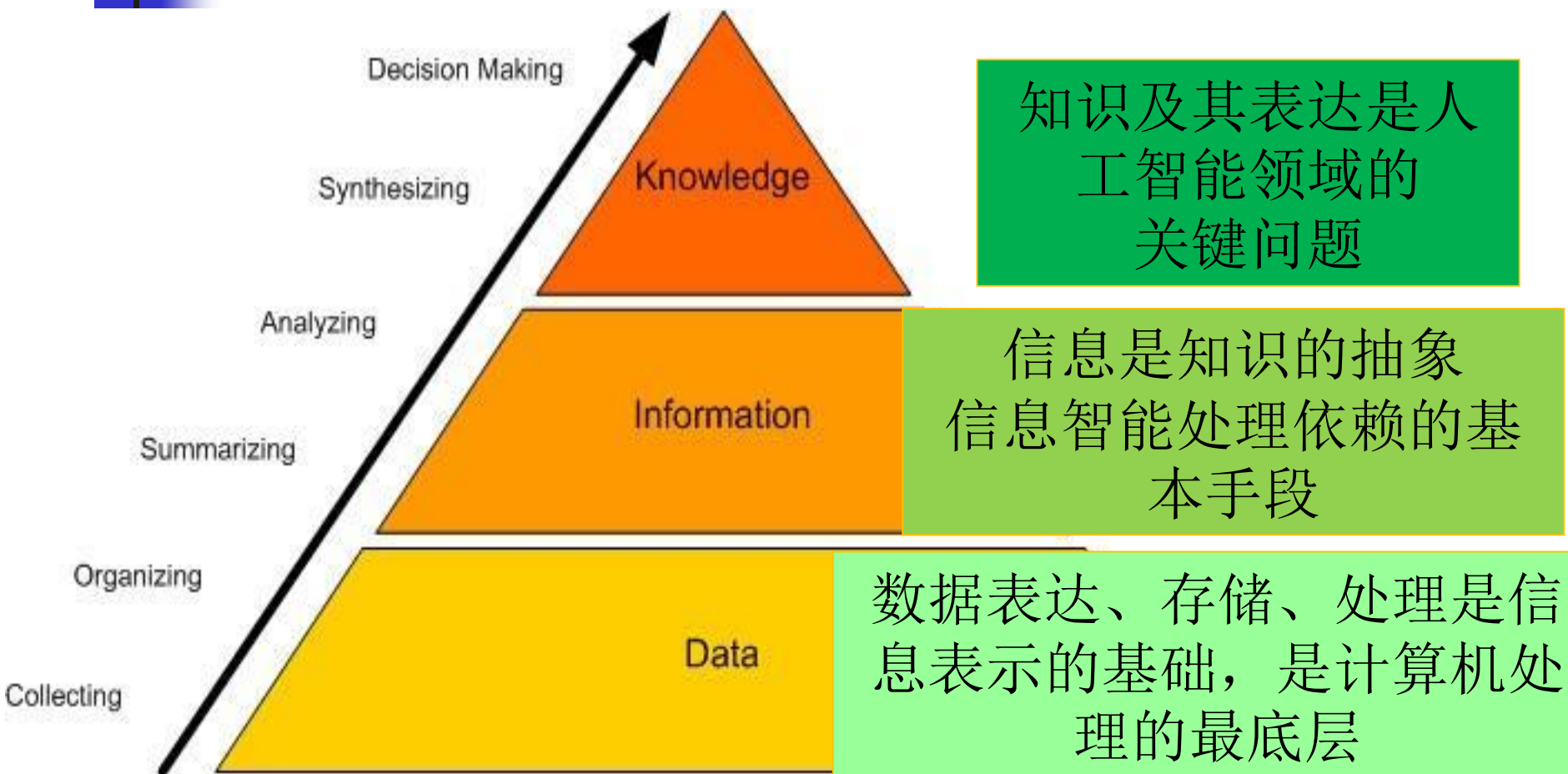


穆铁柱

Google公司在2012年5月提出知识图谱应用与搜索，落地应用：

- 语义搜索：实现Web从网页链接向概念链接转变，支持用户按主题而不是字符搜索
- 关系搜索：获取两个实体之间的关系
- 结构化展示：图形化方式向用户展示经过分类整理的结构化知识

## 1.4 知识表示与信息智能处理





# 从数据到智能：数据模型、信息模型、知识模型和智能模型的描述语言

语义描述能力

