

第一章 数据与信息

- 1、**数据**是现实世界客观事物的符号记录，是信息的载体，是计算机加工的对象。
- 2、在计算机科学中，数据是对所有输入计算机并被计算机识别、存储和处理的符号的总称，是联系现实世界和计算机世界的途径。
- 3、数据的特征：二进制、语义性、分散性、多样性与感知性
- 4、模拟信号是指用连续变化的物理量所表达的信息。其信号的幅度、频率或相位随时间作连续变化，如声音信号、图形信号等。
- 5、数字信号是离散时间信号的数字化表示。其信号的自变量、因变量都是离散的。
- 6、在计算机中，数字信号的大小常用有限位的二进制数表示。
- 7、数字信号的优点：抵抗电路本身干扰和环境干扰的能力强，利于存储、加密与纠错，从而具有较强的保密性和可靠性。
- 8、在现代技术的信号处理中，数据基本上是通过编码将模拟信号转换为数字信号进行存储和传输，文字、图像、声音等类型的数据都可经过编码进行存储和传输。
- 9、文字（字符）编码是效率相对较低的编码方式，有单字节码和双字节码两种。其中，ASCII 码、莫尔斯码属于单字节码，国标码（GBK）、统一码（Unicode）属于双字节码。
- 10、ASCII 码是美国信息交换标准代码，用 8 位二进制码为所有的英文字母（大小写 52 个）、阿拉伯数字（10 个）和常用的不可见控制符（33 个）以及标点符号、运算符等（33 个）建立了转换码，将符号转换为“0”和“1”构成的编码。英文字母 A 和 a 的编码分别为 01000001（十进制数 65）和 01100001（十进制数 97）。
- 11、汉字编码使用的是简体中文的 GB 码和繁体中文的 BIG5 码（大五码）。
- 12、图像编码是指在满足一定保真度的条件下，对图像数据进行变换、编码和压缩，以较少比特数表示图像或图像中所包含的信息的技术。
- 13、位图，最小单位为光栅点（或称像素），因而位图也叫作点阵图（或像素图）。
- 14、在计算机二进制数系统中，每个 0 或 1 就是一个位（bit，数据存储的最小单位），8 个位就称为一个字节（Byte）。即：1 B=8 b 。
- 15、位图文件所占用的空间可按以下公式计算：

文件的大小=文件头+信息头+颜色表项+（图像分辨率*图像量化位数/8）

- （1）文件头：包含文件的类型、大小和位图起始位置等信息，共 14 个字节（B）。
- （2）位图信息头：用于说明位图的尺寸等信息，占 40 个字节（B）。

(3) 颜色信息：用于说明位图中的颜色，有若干个表项，每一个表项定义一种颜色。当图像量化位数为1、4、8时，分别有2、16、256种颜色，每个颜色表项占4个字节；当图像量化位数为24时，没有颜色表项。

(4) 图像分辨率=图像 x 方向的像素数*图像 y 方向的像素数。

(5) 图像量化位数：黑白图像，每一个像素有2种可选颜色（黑、白），称为1位图像。16色图像，每一个像素有16种可选颜色，称为4位图像（ $2^4=16$ ）；256色图像称为8位图像（ $2^8=256$ ）；24位图像的可选颜色更丰富，为 2^{24} 种。

16、声音编码：对声音进行数据编码，必须经过数据的采样、量化和编码。

17、根据奈奎斯特采样定理，如果以一定时间间隔对某个信号 $f(t)$ 进行采样，并且采样频率高于该信号最高频率的两倍，则采样值包含了原信号的全部信息。对于音频信号，常用的采样频率有三种：44.1kHz、22.05kHz 和 11.025kHz。

18、量化是把样值信号的无限多个可能的取值，近似地用有限个数的数值来表示。

19、编码是将量化后的采样值用二进制数码表示，并转换为由二进制编码 0 和 1 组成的数字信号。

20、声音存储空间遵循如下公式：

声音存储空间=采样频率*量化位数*声道数*时间/8（其中立体声声道数为2）

例题：采样频率为44.1kHz、量化位数为16位的立体声，1秒声音所需字节数为： $44.1*1000*16*2*1/8=176400$ （B）。

21、信息是经过加工处理的、具有意义的数据。信息是对客观世界中各种事物的运动状态和变化的反映，是客观事物之间相互联系和相互作用的表征。

22、信息的特征：普遍性、传递性、共享性、依附性和可处理性、时效性、真伪性、价值相对性。

第2章 知识与数字化学习

1、数字化学习工具：Python、思维导图、网络画板。

2、认识自然、探究规律的方法：做实验→获取观察数据→分析处理数据→推理建立数学模型→实验验证模型→形成知识→应用知识解决问题。

3、

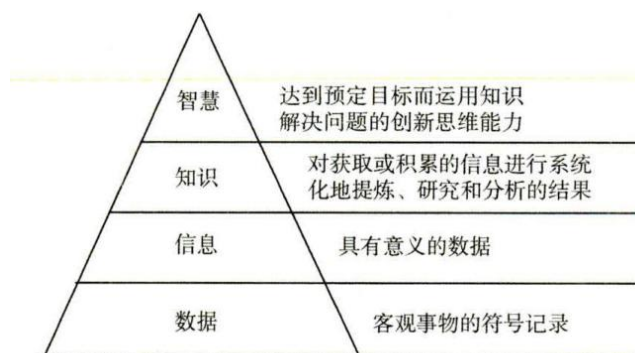


图2-8 数据、信息、知识和智慧的相互关系

4、数字化工具与资源

数字化工具是指能够采集、获取、检索、表示、传输、存储和加工多媒体数字化资源的设备装置。按其功能不同，数字化工具可大致分为以下几种：

(1) 数据采集（模数转换）工具。如声卡、视频卡、数码照相机、数码录像机、手写板、扫描仪等。

(2) 信息获取工具。如台式电脑、笔记本电脑、平板电脑和手机等。

(3) 信息检索工具。如搜索引擎和电子词典等。

(4) 数字化可视化表达工具。如思维导图和演示文稿等。

(5) 信息通信交流与传输工具。如电子邮件、博客、社交软件和视频会议系统等。

(6) 信息存储工具。如光盘、U 盘、硬盘和服务器等。

(7) 信息加工工具。如编程软件、多媒体制作软件、网站制作工具和信息系统制作工具等。

(8) 三维设计工具。如3D 制作软件、3D 打印机、虚拟现实和增强现实制作系统等。

5、数字化工具与资源的优势

(1) 获取的便捷性。 (2) 形式的多样性。 (3) 资源的共享性。

(4) 平台的互动性。 (5) 内容的扩展性

6、数字化学习的特点

数字化学习是伴随计算机多媒体技术、互联网通信技术的发展而产生的，是信息社会的重要特征。与传统的学习方式不同，数字化学习具有问题化、合作性、个性化、创造性和再生性、开放性等特点。

第三章 算法基础

1、编写计算机程序解决问题要经过分析问题、设计算法、编写程序、调试运行程序等若干个步骤。(P46)

2、算法是指在有限步骤内求解某一问题所使用的一组定义明确的规则。通俗地说，算法就是用计算机求解某一问题的方法，是能被机械地执行的动作或指令的有穷集合。(P48)

3、算法的特征：有穷性、确定性、数据输入、数据输出、可行性 (P49)

(1) 有穷性。一个算法在执行有穷步之后必须结束，即一个算法所包含的计算步骤是有限的。

(2) 确定性。算法执行的每一个步骤必须有确切的定义，不能出现模棱两可的情况。

(3) 数据输入。一个算法必须有零个或多个数据输入，以刻画运算对象的初始情况。例如，在上面的算法中，就没有数据输入。

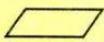
(4) 数据输出。一个算法有一个或多个数据输出，以反映对输入数据加工后的结果，没有输出的算法是毫无意义的。

(5) 可行性。算法中执行的任何计算步骤都可以被分解为基本的可执行的操作步骤，即每个计算步骤都可以在有限时间内完成。

4、描述算法的常用方法有自然语言描述算法、流程图描述算法和伪代码描述算法。(P49-51)

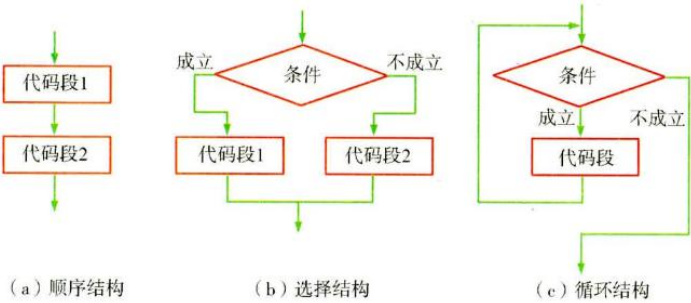
5、用自然语言描述算法，就是用人们日常所用的语言，如汉语、英语等来描述算法。

6、用流程图描述算法是用程序框图来描述算法的一种表示方法

图形	名称	功能
	开始/结束	表示算法的开始或结束。
	输入/输出	表示算法中变量的输入或输出。
	处理	表示算法中变量的计算与赋值。
	判断	表示算法中的条件判断。
	流程线	表示算法中的流向。
	连接点	表示算法中的转接。

7、用伪代码描述算法就是用介于自然语言和计算机语言之间的文字和符号来描述算法。它不用图形符号，书写方便，格式紧凑，易于理解，便于向计算机程序设计语言过渡。

8、三种基本控制结构：顺序结构、选择结构和循环结构 (P52)



9、计算机程序是指为了得到某种结果而可以由计算机等具有信息处理能力的装置执行的代码化指令序列，或者可被自动转换成代码化指令序列的符号化指令序列或者符号化语句序列。简而言之，计算机程序就是指计算机可以识别运行的指令集合。(P54)

10、计算机主要包括运算器、控制器、存储器、输入设备和输出设备五大基本部件

11、计算机程序设计语言，是指一组用来定义计算机程序的语法规则，通常简称为“编程语言”。它是一种被标准化的交流技巧，用于向计算机发出指令。（P54）

12、计算机程序设计语言的发展，经历了从机器语言、汇编语言到高级语言的发展历程。（P55）

13、高级语言的翻译程序有两种类型：编译程序和解释程序。（P58）

第 4 章 程序设计基础

1、常量是指在程序运行过程中其值始终不发生变化的量，通常是固定的数值或字符串。

2、变量是指在程序运行过程中其值可以发生变化的量。在程序设计语言中，变量可以用指定的名字来代表，即变量由变量的“标识符”（又称“名字”）和变量的“内容”（又称“值”）两部分组成。

Python 程序设计语言规定标识符由字母（A~Z，a~z）、数字（0~9）、下划线组成，并且其第一个字符必须是字母或下划线，而不能是数字，如 s1，k，num，pai 等。在标识符中，字母大小写是有区别的，如“value”与“Value”是两个不同的标识符。

3、Python 数据类型

表4-3 Python常用的基本数据类型

数据类型	类型说明
Number (数值)	Python主要支持以下三种不同的数值类型： 整型 (int)，通常被称为整型或整数。在Python 3.x版本中，整型数没有限制大小。 浮点型 (float)。浮点型由整数部分与小数部分组成，浮点型也可以使用科学计数法表示，如$2.5e2 = 2.5 \times 10^2 = 250$。 复数 (complex)。复数由实数部分和虚数部分构成，可以用$a+bj$或者<code>complex(a, b)</code>表示，复数的实部a和虚部b都是浮点型。
String (字符串)	字符串是以单引号或双引号括起来的任意文本，如<code>'abc'</code> <code>"xyz"</code>等。字符串的索引从0开始，-1 为从末尾开始的位置。
List (列表)	列表是写在方括号 “[]” 之间、用逗号分隔开的元素列表。例如，<code>list1=['Good','Run',1997,2000]</code>和<code>list2=[1, 2, 3, 4, 5, 6, 7]</code>。 列表的数据项不需要具有相同的类型。与字符串的索引一样，列表索引从0开始。Python已经内置确定列表的长度以及确定最大和最小的元素的方法。
Tuple (元组)	Python 的元组与列表类似，不同之处在于元组的元素不能修改。元组使用小括号，创建元组只需要在括号中添加元素，并使用逗号隔开。例如，<code>tup1=('Good','Run',1997,2000)</code>和<code>tup2=(1,2,3,4,5)</code>。元组与字符串类似，可以被索引且下标索引从0开始，-1 为从末尾开始的位置，也可以进行截取。
Sets (集合)	集合是一个无序不重复元素的序列。基本功能是进行成员关系测试和删除重复元素。可以使用大括号 { } 或者 <code>set()</code> 函数创建集合。创建一个空集合必须用 <code>set()</code> 而不是 <code>{}</code>，因为 <code>{}</code> 是用来创建一个空字典的。
Dictionary (字典)	字典是无序的对象集合。与列表的区别在于：列表是有序的对象集合。字典当中的元素是通过键来存取的，而不是通过偏移存取。字典用 { } 标识，它是一个无序的“键 (key)：值 (value)” 对集合。键必须使用不可变类型。在同一个字典中，键必须是唯一的。

4、算术运算符与算术表达式

算术运算符包括 “+” “-” “*” “/” “%” “**” 和 “//”，分别表示加、减、乘、除、求余、幂运算和整除。

5. 逻辑运算符与逻辑表达式

逻辑运算符包括 "and" (与)、"or" (或)、"not" (非) 共三个。由逻辑运算符连接而成的表达式称为逻辑表达式。逻辑表达式的值为 True 或 False。

表4-4 逻辑表达式及对应运算结果

运算符	逻辑表达式	描述	示例
and	x and y	“与”。如果 x 为 False, x and y 返回 False, 否则它返回 y 的计算值。	$(a$ and $b)$ 返回0。 $(c$ and $b)$ 返回1。
or	x or y	“或”。如果 x 是 True, 它返回 x 的值, 否则它返回 y 的计算值。	$(a$ or $b)$ 返回1。 $(c$ or $b)$ 返回2。
not	not x	“非”。如果 x 为 True, 返回 False; 如果 x 为 False, 它返回 True。	not(a and $b)$ 返回 True。

6、关系运算符与关系表达式

判断数据大小关系的运算符称为关系运算符, 关系运算符有">"(大于)、">=" (大于等于)、"<"(小于)、"<=" (小于等于)、"==" (等于)、"!=" (不等于)。

7、赋值运算符与赋值表达式

在 Python 中对变量的赋值通过赋值运算符"="来完成。

表4-5 Python语言的赋值运算符

赋值运算符	描述	示例	赋值运算符	描述	示例
=	直接赋值。	$x=y$ 。	/=	除法赋值。	$x/=y$ 相当于 $x=x/y$ 。
+=	加法赋值。	$x+=y$ 相当于 $x=x+y$ 。	%=	求余赋值。	$x\%=y$ 相当于 $x=x\%y$ 。
-=	减法赋值。	$x-=y$ 相当于 $x=x-y$ 。	**=	指数幂赋值。	$x**=y$ 相当于 $x=x**y$ 。
=	乘法赋值。	$x=y$ 相当于 $x=x*y$ 。	//=	整除赋值。	$x//=y$ 相当于 $x=x//y$ 。

8、成员运算符

表4-6 Python语言的成员运算符

成员运算符	描述	示例
in	x in y 。如果 x 是序列 y 的成员, 则计算结果为 True, 否则为 False。	3 in [1,2,3,4], 计算结果为 True。 5 in [1,2,3,4], 计算结果为 False。
not in	x not in y 。如果 x 不是序列 y 的成员, 则计算结果为 True, 否则为 False。	3 not in [1,2,3,4], 计算结果为 False。 5 not in [1,2,3,4], 计算结果为 True。

9、运算符的优先级

表4-7 Python常用运算符的优先级

优先级	运算符	功能及说明	优先级	运算符	功能及说明
1	**	指数运算。	6	=、%=、/=、//= -=、+=、*=、**=	赋值运算。
2	*, /, %, //	乘、除、求余、整除。	7	is、is not	标识运算符。
3	+, -	加运算、减运算。	8	in、not in	成员运算符。
4	>, >=	大于、大于等于。	9	not	逻辑非运算。
	<, <=	小于、小于等于。	10	and	逻辑与运算。
5	==, !=	等于、不等于。	11	or	逻辑或运算。

10、Python 的函数

表4-8 Python常用函数

函数名称	使用说明
abs(x)	返回x的绝对值。
len(x)	计算变量x的长度，但x必须是可计算长度的类型。
min(x)	返回变量x的最小值。
max(x)	返回变量x的最大值。
int(x)	把变量x转换成整型。
float(x)	将变量x转换成浮点型。
sum(x)	计算变量x中元素的总和。
bin(x)	把数值x转换为二进制数字。
bool(x)	x如果是False、None或是空值就返回False。
chr(x)	返回整数x对应的ASCII字符。
divmod(a,b)	返回a/b的商和余数，以元组的方式返回。
all(x)	x中的所有元素都是True才会返回True。
any(x)	x中的所有元素中只要有一个是True就会返回True。
input(x)	显示出信息x，并要求输入数据。
print()	输出函数。

11、数据的输入格式为：<变量>=input('提示信息：')

12、数据的输出格式为：print([object,...] [, sep=""] [, end='\n'])

说明：

- (1) 方括号中的项是可选的，可以省略，如省略则取系统的默认值。
- (2) object 是要输出的对象，可以是常量、变量或表达式等。

(3) `sep` 后面的空格（可以指定为其他字符）表示每个输出对象之间的分隔符，如果缺省的话，默认值是一个单个的空格。

例如，"`print(4, 5, 6, sep=***)`"，则输出结果为"`4**5**6`"。

(4) `end` 后面的字符串含义为输出文本尾的一个字符串，如果缺省的话，默认值是一个 `\n` 换行符。如果设为其他字符，如 `end=''`，则输出当前行的所有内容后，在末尾加一个空格，不换行接着输出下一个 `print()` 的输出对象。

例如，`print()` 输出示例程序如下：

```
print("hello"end="")
print("world")
```

输出为"hello world"

(5) `print()` 函数支持参数格式化，与 C 语言的 `printf` 类似。

13、顺序结构的应用

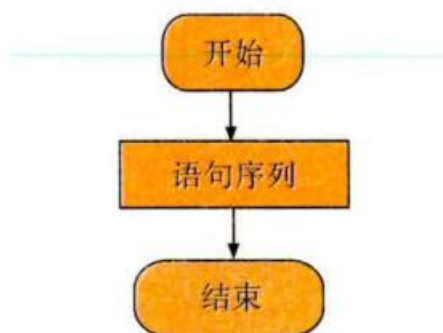


图4-5 顺序结构的流程图

14、选择结构的应用

(1) 单分支 `if` 语句格式：

```
if (条件表达式):
    语句A
```



图4-7 if语句的执行过程

(2) 双分支 `if` 语句格式：

```
if (条件表达式):
    语句A
else:
    语句B
```

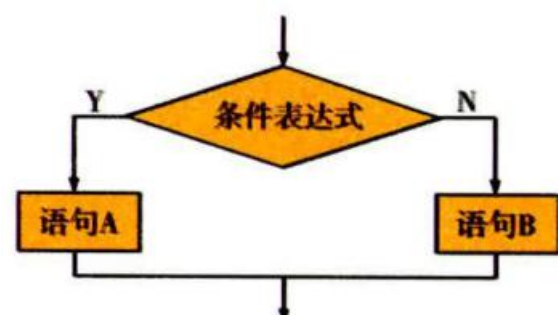


图4-9 双分支if语句执行过程

(3) 多分支 if...elif...else 语句格式:

```
if (表达式1):
    语句1
elif (表达式2):
    语句2
.....
elif (表达式N):
    语句N
else:
    语句N+1
```

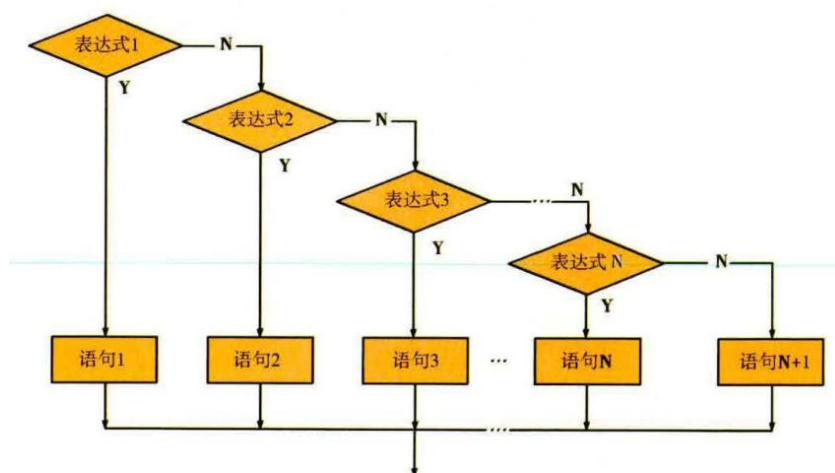


图4-10 if...elif...else语句执行过程

15、for 循环语句的格式:

```
for 循环变量 in 列表:
    语句或语句组
```

```
for value in [1,2,3,4,5]:
    print(value)
    print('good')
```

```
for value in [1,2,3,4,5]:
    print(value)
    print('good')
```

```
values=[1,2,3,4,5]
for value in values:
    print(value)
    print('good')
```

16、range () 函数是让 Python 从指定的第一个值**开始数**，并在到达所指的第二个值后**停止**，因此不包含**第二个值**。使用 range () 函数时，还可指定**步长**。

示例:

```
for value in range(2,11,2):
    print(value)
```

17、while 循环语句格式:

```
while (表达式):
    语句或语句组
```

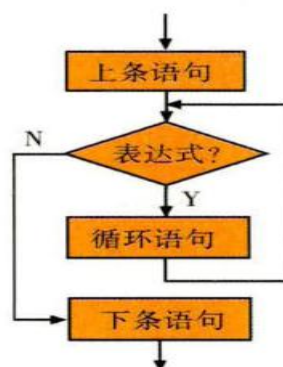


图4-15 while循环执行过程流程图

18、循环控制

(1) **break** 可以跳出当前循环体，从而中断当前循环。

(2) **continue** 不是立即跳出循环体，而是当条件符合需要跳转的时候，跳过该次循环结束前的语句，回到循环开头的条件测试部分，重新开始执行循环。

第 5 章 数据处理与可视化表达

1、数据是指无法在可承受的时间范围内用常规软件工具进行高效**捕捉、管理和处理**的数据集合，是需要**新处理模式**才能具有更强的**决策力、洞察发现力和流程优化能力**的海量、高增长率和多样化的**信息资产**。

2、大数据的产生是与人类日益普及的**网络行为**所伴生的。物联网、云计算、移动互联网、车联网、手机、电脑以及遍布地球各个角落的各种各样的传感器，无一不是数据的来源或是承载的方式。

3、大数据的特征：

(1) 从互联网产生大数据的角度来看，大数据具有“4V”特征：**大量 (Volume)、多样 (Variety)、低价值密度 (Value)、高速 (Velocity)**。

(2) 从互联网思维的角度来看，大数据具有三个特征：**样本渐趋于总体，精确让位于模糊，相关性重于因果**。

(3) 从大数据存储与计算的角度来看，大数据具有两个特征：**分布式存储和分布式并行计算**。

4、传统数据与大数据的区别

传统数据	大数据
数据量小	数据体量巨大
数据类型少	数据类型繁多
价值密度高	价值密度低
更新速度慢	更新速度快
追求数据精确性	追求数据模糊性
本地存储	分布式存储

5、大数据使人们日常生活更为便捷：**方便支付、方便出行、方便购物与产品推介、方便看病与诊病**

6、大数据对人们日常生活产生的负面影响：**个人信息泄露、信息伤害与诈骗**

7、数据采集的基本方法包括系统日志采集法、网络数据采集法和其他数据采集法。

8、在信息系统中，系统日志是记录系统中硬件、软件和系统问题的信息文件。

9、网络数据采集是指通过网络爬虫或网站公开 API（应用程序接口）等方式从网站上获取数据信息。

10、存储数据主要有两种方式，一种是把数据存在本地内部，另一种是把数据放在第三方公共或私有的“云端”存储。

11、数据安全保护技术。数据安全保护指数据不被破坏、更改、泄露或丢失。安装杀毒软件和防火墙只能防备数据安全隐患，而采用拷贝、备份、复制、镜像、持续备份等技术进行数据保护才是更为彻底、有效的方法。

12、解决隐私泄露问题有三个办法：一是技术手段，常用的隐私保护有：①数据收集时进行数据精度处理；②数据共享时进行访问控制；③数据发布时进行人工加扰；④数据分析时进行数据匿名处理等。二是提高自身的保护意识。三是要对数据使用者进行道德和法律上的约束。

12、数据分析一般包括特征探索、关联分析、聚类与分类、建立模型和模型评价等。

（1）特征探索：对数据进行预处理，发现和处理缺失值，异常数据、绘制直方图，观察数据分布的特征，求最大值、最小值、极差等描述性统计量。

（2）关联分析：分析发现存在于大量数据之间的关联性和相关性，从而描述一个事物的共同规律和模式。

（3）聚类分析：是一种探索性的分析。不必事先给出一个分类标准，而是让其自动分类。

（4）数据分类：是数据分析中最基本的方法。先基于样本数据构建分类器，然后进行预测。

第 6 章 人工智能及其应用

1、人工智能是研究计算机模拟人的某些感知能力、思维过程和智能行为（如学习、推理、思考、规划等）的学科。

2、智能问答系统主要包括常见问题解答（FAQ）、问题理解、信息检索、文档库、答案抽取五大模块。

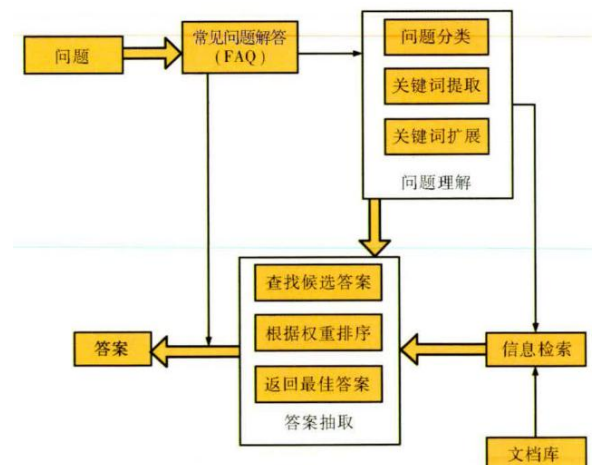


图6-5 典型的智能问答系统结构示意图

3、问题理解模块

该模块主要实现计算机理解用户的问题，确定问题的关键词和问题的类型，为后面的信息检索和答案提供服务。问题理解模块的实现过程一般包括问题预处理、问题分类、关键词提取和关键词扩展等。其中，问题分类主要确定问题的类别，以方便信息检索和答案抽取。问题理解模块主要运用的技术有分词、同义词词典、分类方法等。

4、信息检索模块

该模块主要从互联网或者知识库中找到与问题相关的文档作为答案提取的原材料。信息检索的方法一般有两种，一种是直接利用搜索引擎检索信息，另一种是建立特定的知识库，然后根据知识库建立索引模块，从而可以方便、快速找到相关文档，并根据特点的排序算法对文档进行排序。信息检索模块运用的技术主要包括查询扩展、语料库的构建技术、词汇索引、文档排序等。

5、文档库模块

文档库用于存放专家提供的知识，其内部含有大量某个领域的常识性知识和专家水平的知识与经验总结，且能够利用专家的知识 and 解决问题的方法来处理该领域问题。

6、答案抽取模块

该模块主要利用问题的类型构建相应的答案抽取策略，从信息检索后的文档中对排序靠前的文档进行答案的定位和输出，所用技术主要有答案抽取模板的制定、模式匹配、聚类等。

7、图灵测试是指测试者在与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。问过一些问题后，如果被测试者有超过 **70%** 的答复不能使测试者确认出哪个是人、哪个是机器，那么这台机器就通过了测试，并被认为具有人类智能。

8、人工智能发展大致分为三个阶段。

第一阶段（20 世纪 50-80 年代）刚刚诞生，符号主义快速发展。

第二阶段（20 世纪 80 年代-90 年代末）专家系统快速发展，数学模型有重大突破。

第三阶段（21 世纪初至今）大数据的积聚、理论算法的革新、计算能力的提升，人工智能进入繁荣时期。

9、人工智能在生活中的应用有智能制造、智能家居、智能教育、智能交通、智能安防、智能医疗、智能物流。