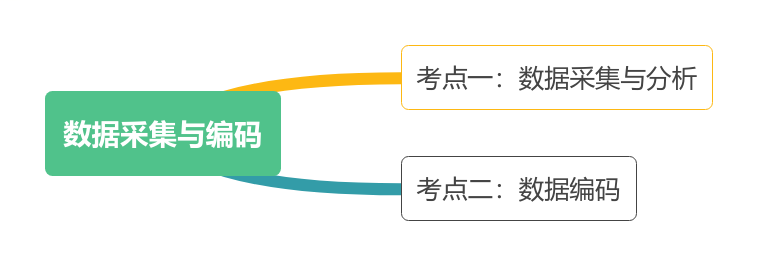
******专题02 数据采集与编码**





**考点一：数据采集与分析**

1．(2023贵州)下数据采集的方法，不可取的是（ ）

A．利用问卷星发放调查问卷采集数据 B．利用木马程序收集数据

C．通过走访调查采集数据 D．通过实验操作采集数据

【答案】B

【知识点】数据的采集

【详解】本题考查常见的数据采集方式。A、C、D选项都是合法且有效的数据采集方法，而B选项（利用木马程序收集数据）则违反了隐私保护和信息安全的原则，且可能触犯法律，因此是不可取的。

因此，答案是B。

2．(2023内蒙古)通过地图、遥感图像、野外考察等获得数据的过程是（ ）

A．数据处理 B．数据输出 C．数据分析 D．数据采集

【答案】D

【知识点】数据的采集

【详解】本题考查的是数据的采集。数据处理通常发生在数据收集之后，涉及对原始数据进行清洗、整理、转换等操作，以便于后续的数据分析。数据输出常指的是将处理后的数据以某种形式（如报告、图表等）呈现出来，供用户或决策者使用。数据分析涉及对处理后的数据进行统计、建模、挖掘等操作，以揭示数据中的模式、趋势或关联。数据采集是通过各种手段和方法收集原始数据的过程，包括观察、实验、调查等。综上所述，通过地图、遥感图像、野外考察等获得数据的过程是数据采集，即D选项。

3．（2024全国）数据按照获取的渠道分为直接数据和间接数据。下列属于直接数据的是（ ）

A．咨询获得的景点介绍 B．查阅县志了解的历史

C．实验得到的电流数值 D．网络搜索的空气指数

【答案】C

【知识点】数据采集

【详解】本题考查的直接数据与间接数据。直接数据：指通过直接观察、实验或测量等手段获取的第一手数据。间接数据：通常是通过查阅文献、资料、报告等途径获得的，这些数据已经经过了加工或整理。

分析选项：

A选项通过咨询获得的，属于二手信息，即间接数据。

B选项通过查阅历史资料获得的，同样是二手信息，即间接数据。

C选项通过直接实验测量得到的，属于第一手数据，即直接数据。

D选项通过网络搜索获得的，通常是基于其他来源的数据整理得到的，属于间接数据。

因此，答案是C。

4．（2023广西）在成都第31届世界大学生夏季运动会上，各赛场安装的高清摄像机，实时记录比赛情况。

摄像机记录信息的过程属于（ ）

A．数据可视化 B．数据分类 C．数据分析 D．数据采集

【答案】D

【知识点】数据的采集

【详解】本题考查的是数据的采集。摄像机实时记录比赛情况，这是一个信息收集和记录的过程。即D选项

5．（2024安徽）某平台根据用户的搜索需求，进行智能推送的过程属于（ ）

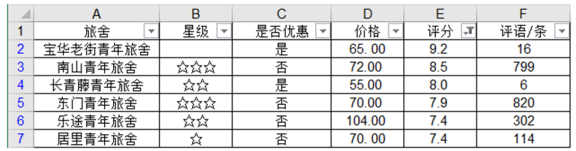
A．数据分析 B．数据分类 C．数据采集 D．数据可视化表达

【答案】A

【知识点】数据分析

【详解】本题考查的是数据分析。某平台根据用户的搜索需求，智能推送相关内容。这涉及到对用户搜索数据的处理，以便理解用户的需求并作出相应的响应。因此，答案是A选项（数据分析）。

6．（2024安徽）如图所示的数据表，筛选星级为“三星”，并且评分大于7.0的旅舍，此过程属于（ ）



A．数据采集 B．数据分析 C．数据存储 D．数据可视化表达

【答案】B

【知识点】数据分析

【详解】本题考查的是数据分析。根据提供的数据表，我们需要筛选出满足特定条件（星级为“三星”，且评分大于7.0）的旅舍。这是一个对数据进行筛选和处理的过程。因此，B选项正确。

7．（2024安徽）某商场年末促销活动中的部分商品客流量变化如图所示，双十二当天各品类客流量均增幅明显，圣诞节、元旦前一天，各品类客流亦出现不同幅度的增长。这里运用的数据分析方法是（ ）



A．对比分析 B．平均分析 C．现实分析 D．聚类分析

【答案】A

【知识点】数据的分析

【详解】本题考查的是数据的分析。根据题目描述的场景和数据分析方法的定义，最符合的数据分析方法是A选项（对比分析）。这种方法通过对比不同时间段的数据，揭示了商场促销活动对客流量的影响。

8．（2024云南）“词云图”是数据可视化的一种表达方式。2023年国家社科基金立项热词的词云图如下图所示，下列说法错误的是（ ）



A．通过词云图显示，“机制研究”比“实现路径”的出现频率高

B．制作词云图一般需要经过数据准备、分词处理、特征提取、数据分析、结果呈现等步骤

C．词云图必须显示该数据集包含的全部词语

D．词云图通常使用文字的大小和颜色来表示关键词出现的频次

【答案】C

【知识点】数据的可视化表达

【详解】本题考查的是数据的可视化表达。词云图是一种数据可视化表达方式，通过文字的大小和颜色来表示关键词出现的频次。在词云图中，通常出现频率越高的词，其字体越大，颜色也可能越醒目。

词云图通常根据关键词的重要性（如出现频率）来选择性地显示词汇，而不是必须显示全部词语。如果数据集包含大量词汇，词云图可能会过滤掉一些出现频率较低或不重要的词汇，以便更清晰地展示关键信息。因此，C选项的说法是错误的。

9．（2023贵州）为分析学生成绩，张老师将考试成绩做成一张柱状图呈现给学生，请问张老师所做的是数据处理哪一步（ ）

A．数据收集 B．数据整理 C．数据分析 D．数据可视化

【答案】D

【知识点】数据的可视化表达

【详解】本题考查的是数据的可视化表达。张老师将考试成绩做成柱状图，这是一个将数据以图形形式展示出来的过程。张老师所做的是数据处理的“数据可视化”步骤。

10．（2024江苏）大数据个人报表

近年来，很多网络平台会在年底为用户生成专属的个人报表，显示一年内该用户在该平台上的各种使用行为数据。

（1）这类个人报表是利用大数据技术收集用户的个人行为数据，并通过分类和计算获得。由此可见( )

A.大数据采集的数据必须准确无误

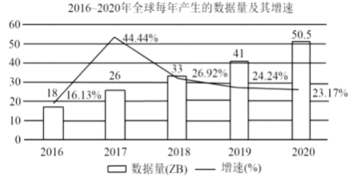
B.大数据的数据采集主要还是人工方式

C.大数据分析的是抽样数据

D.大数据分析能发现数据间的相关性

（2）国际数据公司统计了2016-2020年全球每年产生的数据量及其增速，所得结果如图所示。

此统计主要采用的数据分析方法是 ，使用了 和 两种图表来组合进行可视化呈现( )



A.聚类分析、饼图、条形图 B.对比分析、折线图、柱形图

C.结构分析、饼图、条形图 D.平均分析、折线图、柱形图

（3）该网站可实现对数据的加密解密，应用界面如图。可见该应用的服务架构体系属于( )



A.B/S B.P2P C.C/S D.B/C

【答案】（1）D，（2）B，（3）A

【知识点】数据的分析和可视化表达

【详解】本题考查的是数据的分析和数据的可视化表达。

（1）大数据采集的数据不一定完全准确无误，A选项错误；大数据的数据采集主要是通过自动化的方式，而非人工方式，B选项错误；大数据分析的是全体数据，而非抽样数据，C选项错误；通过对用户个人行为数据的收集和分析,能够发现不同数据之间的相关性,生成个人报表，D选项正确。故答案为:D。

（2）国际数据公司统计不同年份的数据量及其增速，主要是进行对比分析，以观察数据的变化趋势和差异。使用了折线图来展示数据量的增速变化趋势，使用柱形图来直观呈现每年产生的数据量的具体数值。故答案为:B。

（3）B/S架构是指浏览器/服务器架构，用户通过浏览器访问服务器端的应用程序。从图中给出的浏览器界面可以判断该应用的服务架构体系属于B/S架构。故答案为:A。

**考点二：数据编码**

1．（2024江苏）以下说法正确的是（ ）

|  |
| --- |
| A．用传感器直接获得的信号一定是模拟信号 |
| B．数字信号随时间的变化是非连续的 |
| C．数字信号无法由模拟信号转换得来 |
| D．模拟信号的值可以直接用计算机所能理解的二进制表示 |

【答案】B

【知识点】模拟信号与数字信号

【详解】本题考查的是模拟信号与数字信号的相关知识。传感器可以获取各种物理量（如温度、压力、光强等）并将其转换为电信号。这些电信号可以是模拟信号，也可以是数字信号，这取决于传感器的类型和设计，所以A选项错误。数字信号是由离散的数值（通常是二进制数）组成的，这些数值在时间上是不连续的，即它们只在特定的时间点（如采样时刻）上取值，因此，B选项正确。模拟信号可以通过采样、量化和编码等过程转换为数字信号，因此，C选项错误。计算机使用二进制进行数据处理，但模拟信号的值通常需要先进行采样、量化和编码才能转换为计算机可以理解的二进制形式。直接表示是不可能的，因为模拟信号的值是连续的，而计算机处理的是离散的二进制数。因此，D选项错误。

本题答案选B

2．（2023江苏）英语老师上课播放的英语听力，可以在电脑中播放音频文件，也可以利用传统收录机播放磁带。这两种声音信号分别是（ ）

A．数字信号、数字信号 B．模拟信号、模拟信号

C．数字信号、模拟信号 D．模拟信号、数字信号

【答案】C

【知识点】模拟信号与数字信号

【详解】本题考查的是模拟信号与数字信号相关的知识，英语老师上课播放的英语听力，在电脑中播放的是数字信号，而利用传统收录机播放的磁带则是模拟信号。故选C。

本题答案选C

3．（2022福建）将模拟信号转换为数字信号的过程包含采样和（ ）

A．解密、量化 B．压缩、编码 C．量化、编码 D．压缩、量化

【答案】C

【知识点】模拟信号与数字信号的转化

【详解】本题考查的是模拟信号与数字信号的转化。将模拟信号转换为数字信号的过程包含采样、量化和编码。因此，正确答案是C。

本题答案选C

4．（2023内蒙古）有4盏信号灯，信号灯只有“开”和“关”两种状态，如果包括4盏信号灯全关的状态，则最多能表示的信号编码数为（ ）

A．10 B．5 C．20 D．16

【答案】D

【知识点】数据编码

【详解】本题考查的是二进制编码，

第一盏信号灯有2种状态：开或关。

第二盏信号灯也有2种状态：开或关。

第三盏信号灯还有2种状态：开或关。

第四盏信号灯同样有2种状态：开或关。

那么，4盏信号灯的所有可能组合就是：16种

这16种组合包括了所有信号灯都关的状态，也包括了其他所有可能的开关组合。

所以，最多能表示的信号编码数是16

5．（2024安徽）已知字母B的ASCII码二进制表示为01000010，则字母D的ASCII码值用十进制表示为（ ）

A．66 B．67 C．68 D．69

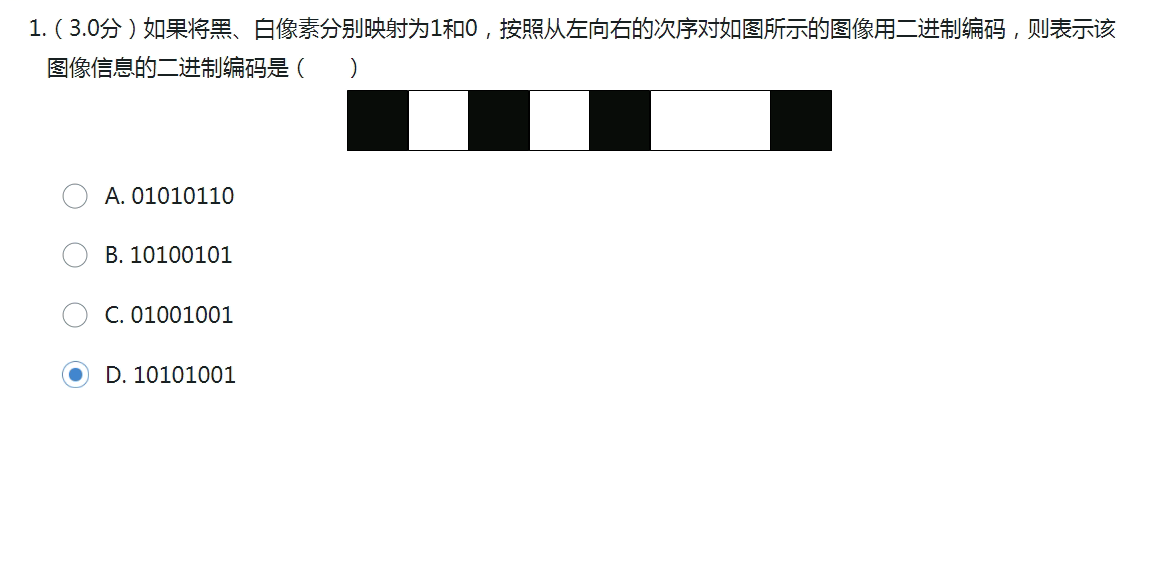
【答案】C

【知识点】ASCII码值的计算

【详解】本题考查的是ASCII码值计算，注意到ASCII码表中，字母是顺序排列的。字母B的ASCII码二进制表示为01000010，转化成十进制是66。因此，字母D的ASCII码值用十进制表示为68，选项C是正确的。

本题答案选C

6．（2023安徽）如果将黑、白像素分别映射为1和0,按照从左向右的次序对如图所示的图像用二进制编码，则表示该图像信息的二进制编码是： （ ）



A.01010110 B.10100101

C.01001001 D.10101001

【答案】D

【知识点】图像编码

【详解】本题考查的是计算机中图像编码。黑像素对应1，白像素对应0，按照从左向右的次序对图像进行二进制编码10101001。故选D。

本题答案选D

7．（2023全国）交通控制中心对信号灯状态进行数据编码，如图1所示。若只允许左转一个车道通行，对应的编码是110101，如图2所示。若不允许车辆左转，只允许车辆直行和右转，对应的编码是（ ）

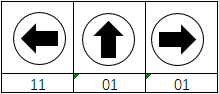
 

图1 图2

A．010111 B．111111 C．011111 D．100101

【答案】C

【知识点】数据编码

【详解】本题考查的是数据编码。不允许左转，左转为红灯01，允许直行和右转，直行和右转为均绿灯11，所以编码为011111。故选C。

8．（2024湖北）曹雪芹的《红楼梦》前80回大约有61万字，加上高鹗续写的40回共有约73万字。一个汉字存储时占用2个字节，如果要把《红楼梦》全书做成电子书（加注标点符号），电子书的大小约为（ ）

A．1.5MB B．700MB C．100MB D．5KB

【答案】A

【知识点】数据编码

【详解】本题考查的是字符和数据编码。根据提供的问题，我们需要计算《红楼梦》全书做成电子书后的大小。这个问题与图片中的浏览器窗口内容、网络地址、按钮或文字文本无关。已知《红楼梦》全书共有约73万字，一个汉字存储时占用2个字节。因此，我们可以计算出全书所需的字节数为：730,000字 × 2字节/字 = 1,460,000字节。为了将字节数转换为更常用的存储单位MB，使用以下公式：1MB = 1,024KB = 1,024 × 1,024字节。所以，全书的大小约为：1,460,000字节 ÷ (1,024 × 1,024字节/MB) ≈ 1.39MB。

但是，考虑到电子书通常会加注标点符号，这些标点符号也会占用一定的存储空间。因此，我们可以合理推测，加上标点符号后的电子书大小会略大于1.39MB，但远小于其他选项（700MB、100MB）且明显大于5KB。电子书《红楼梦》全书（加注标点符号）的大小最接近1.5MB，所以正确答案是A。

9．（2024江苏）一段音频的参数为“44.1KHz，32位，单声道，8秒”，该音频的采样频率为（ ）

A．44.1KHz B．32位 C．单声道 D．8秒

【答案】A

【知识点】数据编码

【详解】本题考查的是声音编码。采样频率指的是每秒对音频信号进行采样的次数，它决定了音频的清晰度和频率范围。在这个问题中，采样频率被明确给出为“44.1KHz”。