



多媒体技术

2.1.1 常见的媒体元素

- **Unicode标准**
 - **UCS**: ISO/IEC开发的单一字符集
 - **Unicode**: Unicode联盟开发的单一字符集
 - 世界不需要两个不兼容的字符集
 - 从**Unicode 2.0**开始保持两个标准码表的兼容性, 共同调整未来的扩展工作。

2.1.1 常见的媒体元素

- Unicode标准

- 编码空间：17个平面(平面0, ..., 16)

- 用5位(bit)表示平面的编号

- 每个平面包含256行 × 256码位/行=65536个码位，用十六进制数表示，用于表达有含义的抽象字符

- 在理论上说，编码空间可容纳 $17 \times 256 \times 256 = 114\ 112$ 字符，编码空间的范围为0 ~ 0x10FFFF

2.1.1 常见的媒体元素

- UTF编码

- 在Unicode标准中，表示字符位置的码点(编号)长度是**固定的**
- 为节省文件的存储**空间**，尤其是以7位**ASCII**字符为主的西文，并考虑到要与先前的字符编码**兼容**:

Unicode定义了1字节(8位)、2字节(16位)和4字节(32位)的三种Unicode转换格式(Unicode Translation Format或Universal Character Set Transformation Format, UTF)，分别称为**UTF-8**、**UTF-16**和**UTF-32**编码，用来表示一个字符的代码

2.1.1 常见的媒体元素

- **UTF-8 (8-bit UCS/Unicode Transformation Format)**
 - 代码长度可变的转换格式，也是ISO/IEC 10646字符的转换格式；
 - 字符编码用由8位构成的1、2、3或4个字节(码元)表示；
 - UTF-8可表示Unicode标准中的任何字符

2.1.1 常见的媒体元素

- **UTF-8 (8-bit UCS/Unicode Transformation Format)**
 - 1字节码: 用于**U+0000 ~ U+007F**的字符, 对ASCII字符集中的128个字符进行编码, 可使原来处理ASCII字符的软件无需或只做少量修改就可继续使用
 - 2字节码: 用于**U+0080 ~ U+07FF**的1920个字符, 对拉丁文、希腊文、西里尔字母、亚美尼亚语、希伯来文和阿拉伯文等文种的字符进行编码
 - 3字节码: 用于**U+000800 ~ U+00D7FF**和**U+00E000 ~ U+00FFFF**的61 440个字符, 对大部分常用字符, 包括**CJK统一汉字**进行编码
 - 4字节码: 用于**U+010000 ~ U+10FFFF**的1048576个字符进行编码

2.1.1 常见的媒体元素

表 2-18 用 UTF-8 转换 Unicode 字符编码的方法^①

Unicode 码点	UTF-8 编码			
	字节 1	字节 2	字节 3	字节 4
0000~007F (128 个代码)	0xxxxxxx (ASCII 字符集, 最高位为 0)			
0080~07FF (1920 个代码)	110yyyyx (110 开始, 转换码为 2 字节)	10xxxxxx (10 开始)		
0800~FFFF (61440 个代码)	1110yyyy (1110 开始, 转换码为 3 字节)	10yyyyxx (10 开始)	10xxxxxx (10 开始)	
10000~10FFFF (1048576 个代码)	11110zzz (11110 开始, 转换码为 4 字节)	10zzyyyy (10 开始)	10yyyyxx (10 开始)	10xxxxxx (10 开始)

编码规则：字节1的**高位**使用固定的**0、110、1110或11110**，分别表示转换码的长度为1、2、3或4个字节。这样就可根据字节1高位的数值，判断UTF-8转换码的Unicode字符

2.1.1 常见的媒体元素

【例 2.3】根据 UTF-8 的编码规则，对字符 A、©、汉、博、岱进行转换，结果如表 2-19 所示。

表 2-19 部分字符编码

字符	Unicode	Bin	UTF-8 (Bin)	UTF-8 (H)
A	0041	01000001	01000001	41
©	00A9	10101001	11000010 10101001	C2A9
汉	6C49	01101100 01001001	11100110 10110001 10001001	E6B189
博	535A	01010011 01011010	11100101 10001101 10011010	E58D9A
岱	20C30	0 0010 0000 1100 0011 0000	11110000 10100000 10110000 10110000	F0A0B0B0

2.1.1 常见的媒体元素

- **UTF-16 (16-bit UCS/Unicode Transformation Format)**
 - **代码长度可变**的转换格式：可用1个字(2个字节)或2个字(4个字节)表示
 - 对于在Unicode的基本多文种平面(**BMP**)上定义的字符编码，如拉丁字母、CJK中日韩字符、其他文字或符号，用**1个字**表示；对于在**增补平面**上的字符，用**2个字**表示
 - 除U+D800-U+DFFF不是字符外，在U+0000-U+10FFFF之间的所有字符编码都用UTF-16转换

2.1.1 常见的媒体元素

表 2-20 用 UTF-16 转换 Unicode 字符编码的方法

16 进制编码范围	UTF-16 转换码二进制的表示法	10 进制编码范围	字节数
0000 0000~0000 D7FF 0000 E000~0000 FFFF	xxxxxxxx xxxxxxxx	0~65535	2
0001 0000~0010 FFFF	110110yyyyyyyyyy 110111xxxxxxxxxx	65536~1114111	4

编码规则:

如字符编码小于 **0x10000** (0 ~ 65535), 直接用**两字节**表示;
其余用**20位**(bit)表示, 在其前10位的前面加110110, 在其
后10位的前面加110111, 生成**4个字节**的转换码

2.1.1 常见的媒体元素

- UTF-16 (16-bit UCS/Unicode Transformation Format)
 - BMP中代理区(D800-DFFF), 分为前后两个区域
 - **D800-DBFF**: 高半代理区 4行*256=1024个码点
 - **DC00-DFFF**: 低半代理区 4行*256=1024个码点
 - 代理对(Surrogate pair)
 - 从两个区域各取一个码元(16位), 形成“代理对”, 用来表示增补平面码点
 - 所有大于0xFFFF的码点值(即增补平面码点编号, 范围为0x10000~0x10FFFF, 十进制为65536~1114111)要编码成UTF-16编码方式, 必须使用代理对来表示

2.1.1 常见的媒体元素

- UTF-16 (16-bit UCS/Unicode Transformation Format)
 - 增补平面的第一个码点的编号0x10000

其UTF-16编码为： 0xD800 0xDC00

====代理码元1==== ====代理码元2====
1101 10pp ppxx xxxx **1101 11**xx xxxx xxxx

- pppp: 表示16个增补平面
- xx..xx: 增补平面的某个码点
- 共: $1024*1024=1048576$ 个码点, 增补平面 $65536*16$

2.1.1 常见的媒体元素

- UTF-16 (16-bit UCS/Unicode Transformation Format)
 - 例：码点值为10437 (字符名称为U+10437)的字符的UTF-16编码

0x10437减去0x10000，结果为0x00437，二进制为**0000 0000 0100 0011 0111**

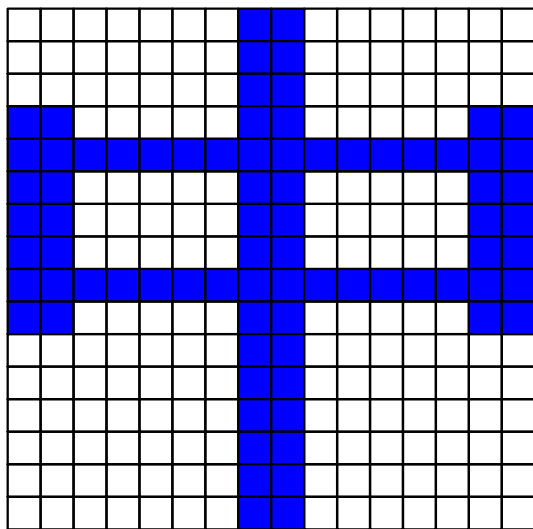
1101 1000 0000 0001 1101 1100 0011 0111

2.1.1 常见的媒体元素

- 字符显示

- 点阵字符

- 字符形状用点阵表示
 - 点阵是由点组成的二维阵列，用来生成字符或图形，大小没有限定



二进制数	十六进制数
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180
110000001100000011	C183
1111111111111111	FFFF
110000001100000011	C183
110000001100000011	C183
110000001100000011	C183
1111111111111111	FFFF
110000001100000011	C183
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180
000000001100000000	0180

要显示“中”，用16*16点阵表示，需要32个字节

2.1.1 常见的媒体元素

- 字符显示

- 点阵字符

- 用于表示字形的阵列称为**字模**，或称点阵字模

优点：显示速度快；

缺点：耗费存储空间；放大后在文字边缘出现锯齿状

- 字模的**点数越多**，显示的汉字越美观
 - 每个字符都有字模，构成字模库，字模与字符的**内码**之间建立**对应关系**，当知道字符的**内码**时，获得字符的**字模**并送到输出设备上显示
 - 1：显示字符颜色；0：不显示，背景颜色

2.1.1 常见的媒体元素

- 字符显示

- 矢量字符

- 字符看成是一个图形，用数学描述构成字符的点、线、弧等及它们的位置，创建的图称为矢量图
 - 矢量汉字组成的字库称为矢量汉字字库，保存的汉字是对每一个汉字如何构造的一组描述
 - 如笔画的起点和终点坐标、半径大小、弧度多少等
 - 在显示或打印矢量字符时，经过一系列的数学运算才能产生字符
 - Windows\Fonts
 - 扩展名FON，表示该文件中的字库是点阵字库
 - 扩展名TTF，表示该文件中的字库是矢量字库，如幼圆

2.1.1 常见的媒体元素

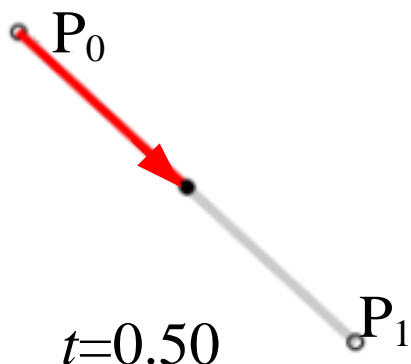
- 贝塞尔曲线
 - 贝塞尔：法国工程师Pierre étienne **Bézier** (1910—1999)的中译名
 - 贝塞尔曲线(Bézier curve): 用数学方法计算的曲线, 用来把多个点连成自由形态的光滑曲线或曲面
 - 只需少量的点就可定义大量的形状, 因此用于逼近一个给定形状时比其他数学方法更有效

2.1.1 常见的媒体元素

- 线性贝塞尔曲线

- 假设有两个点 P_0 和 P_1 ，线性贝塞尔曲线(linear Bézier curve)是用函数 $B(t)$ 描绘的一条直线，在数学上用下式表示：

$$B(t) = P_0 + (P_1 - P_0)t = (1-t)P_0 + P_1t, 0 \leq t \leq 1$$



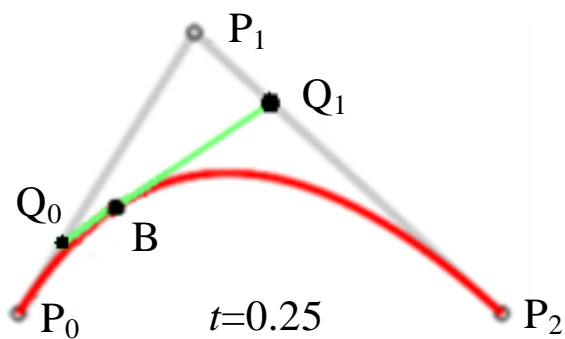
t 用于描述从 P_0 到 P_1 方向走过的路径
 $t=0.5$ 时 $B(t)$ 从 P_0 向 P_1 走了一半的路程
当 t 从0到1时，描绘的曲线是一条 P_0 到 P_1 的直线

2.1.1 常见的媒体元素

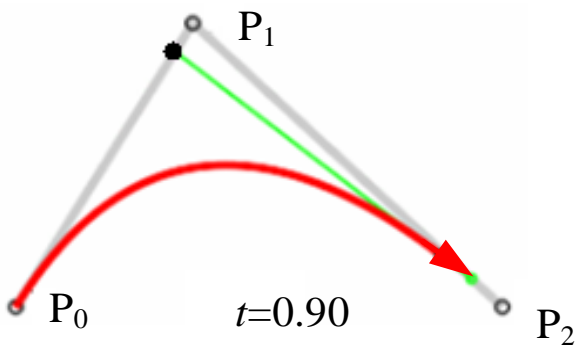
- 二次贝塞尔曲线

- 假设有三个点， P_0 、 P_1 和 P_2 ，二次贝塞尔曲线 (quadratic Bézier curve) 是用函数 $B(t)$ 描绘的路径，在数学上用下式表示：

$$B(t) = (1-t)^2 P_0 + 2(1-t)tP_1 + t^2 P_2, \quad 0 \leq t \leq 1$$



(a) 二次贝塞尔曲线的构造



(b) 二次贝塞尔曲线的生成

2.1.1 常见的媒体元素

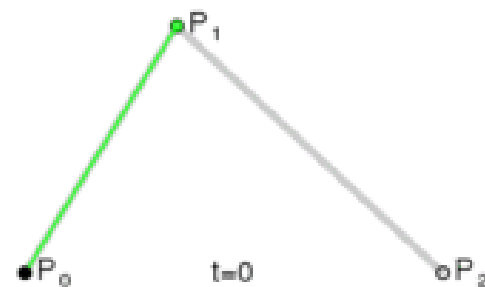
- 二次贝塞尔曲线

- 假设有三个点， P_0 、 P_1 和 P_2 ，二次贝塞尔曲线 (quadratic Bézier curve) 是用函数 $B(t)$ 描绘的路径，在数学上用下式表示：

$$B(t) = (1-t)^2 P_0 + 2(1-t)tP_1 + t^2 P_2, \quad 0 \leq t \leq 1$$

当 t 从 0 变到 1 时：

- ✓ 通过 P_0 和 P_1 计算出点 Q_0 ， Q_0 从 P_0 变化到 P_1 的过程中，描绘一条线性贝塞尔曲线
- ✓ 通过 P_1 和 P_2 计算出点 Q_1 ， Q_1 从 P_1 变化到 P_2 的过程中，描绘一条线性贝塞尔曲线
- ✓ 通过 Q_0 和 Q_1 计算出点 B



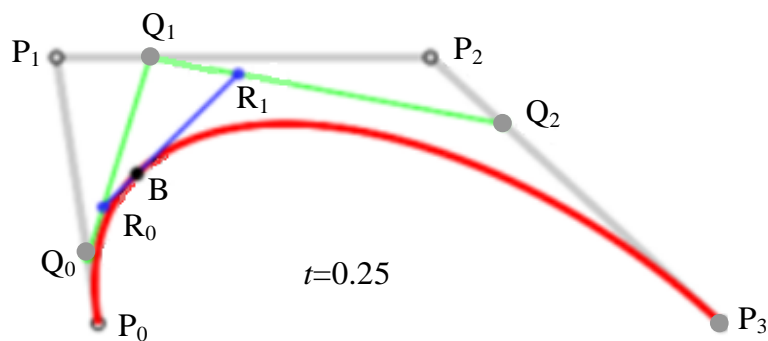
公式推导

2.1.1 常见的媒体元素

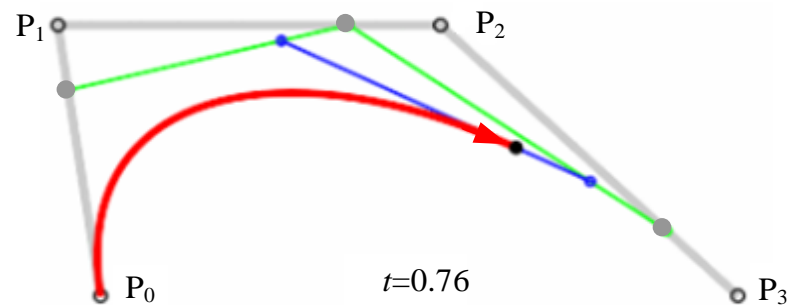
- 三次贝塞尔曲线

- 假设在三维空间的平面上有四个点， P_0 、 P_1 、 P_2 和 P_3 ， P_0 到 P_3 的三次贝塞尔曲线(cubic Bézier curve)用函数 $B(t)$ 描绘，在数学上用下式表示：

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t) t^2 P_2 + t^3 P_3, \quad 0 \leq t \leq 1$$



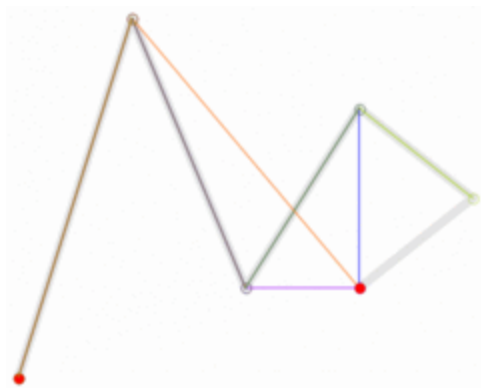
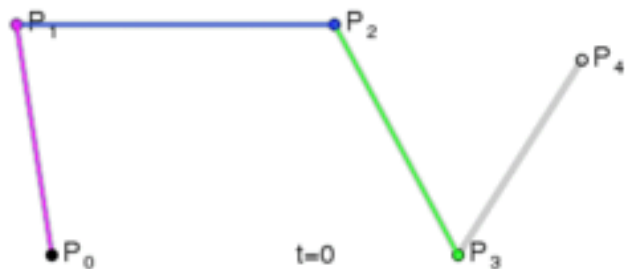
(a) 三次贝塞尔曲线构造



(b) 三次贝塞尔曲线的生成

2.1.1 常见的媒体元素

- 三次贝塞尔曲线



2.1.1 常见的媒体元素

- 文本文件分类
 - 非格式化文本文件
 - 只有文字信息，没有其他格式信息
 - 格式化文本文件
 - 带有各种文本排版信息等格式信息的文本文件
 - 段落格式、字体格式、文章的编号、分栏、边框
 - 文本编辑工具
 - Office、WPS、Editplus、Notepad++

2.1.1 常见的媒体元素

- 数据可视化

- 将文本数据以**视觉**的形式来**呈现**，如图表或地图等，以帮助人们了解这些数据的意义

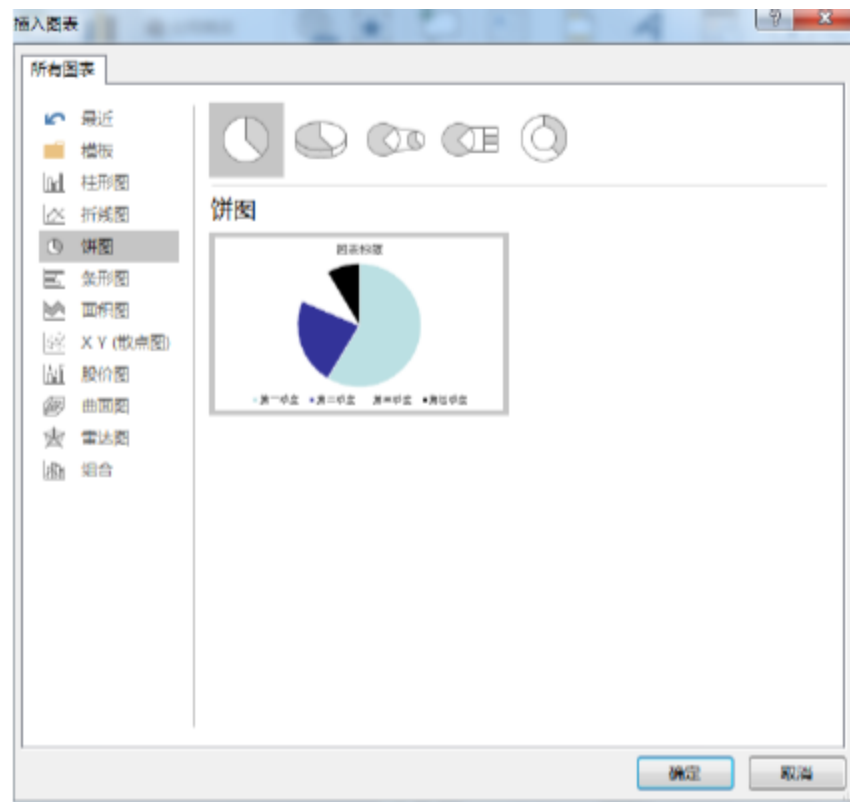
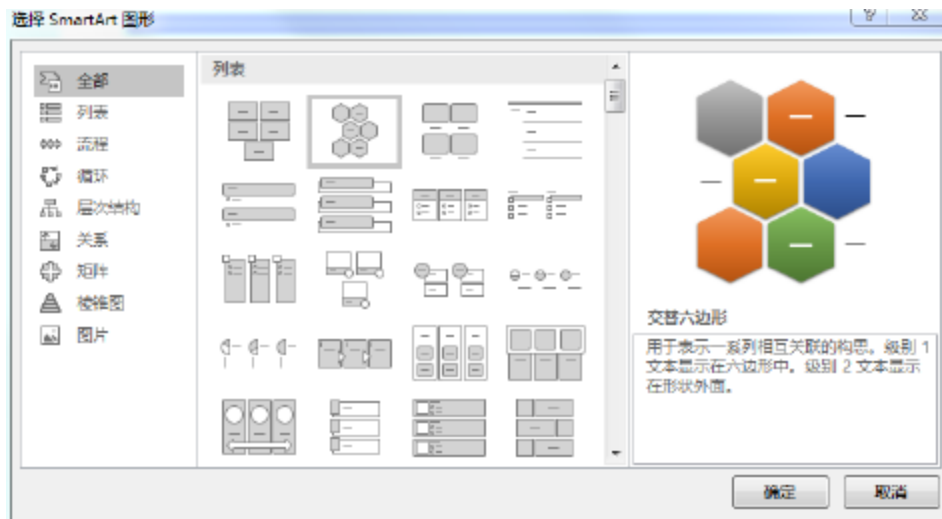


全球24小时航班流量图

多 媒体 技术

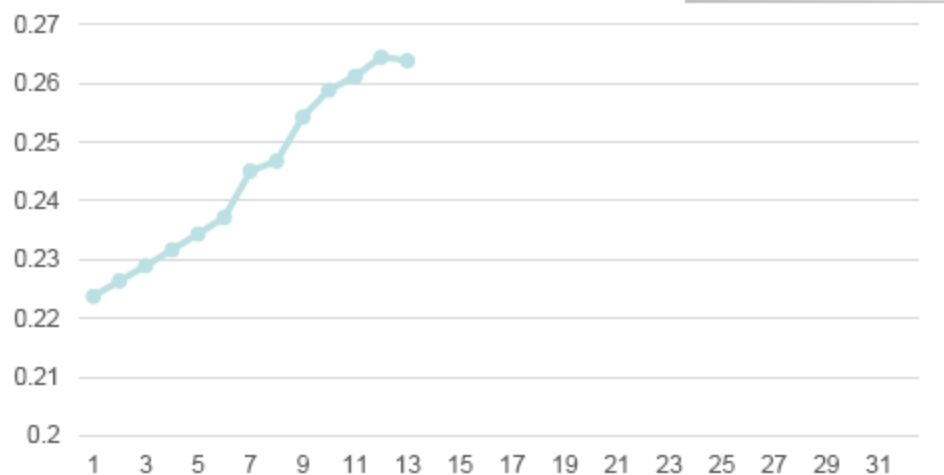
2.1.1 常见的媒体元素

- 数据可视化
 - PowerPoint



2.1.1 常见的媒体元素

- 数据可视化
 - Excel

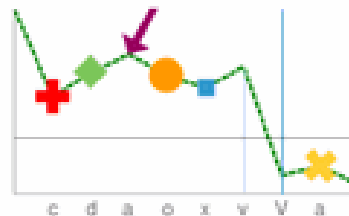
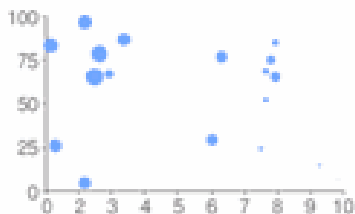
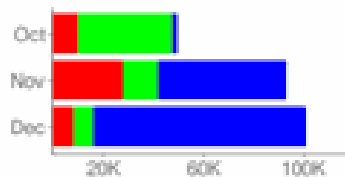
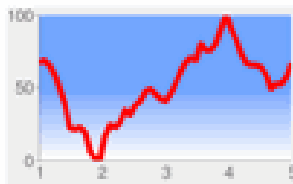


2.1.1 常见的媒体元素

- 数据可视化

- Google charts

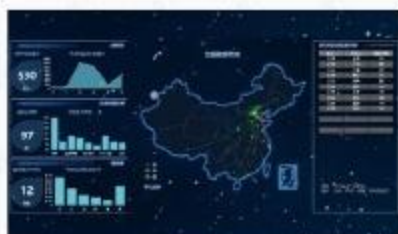
- Google chart API, 制图服务接口, 用来为统计数据自动生成图片
 - 用于在web上可视化数据, 可在web页面上嵌入 javascript



2.1.1 常见的媒体元素

- 数据可视化
 - SmartBi

<http://demo.smartbi.com.cn/smartbi/vision/index.jsp>



航线分析



地区民生



热销车销售分析



汽车大屏



区域销售情况



电网案例



地图分析



销售情况

作业2

- 使用**HTML5 canvas**的贝塞尔曲线方法绘制学号的最后一位。（**IE可运行**，曲线光滑更好）
- 下周上课前将代码（文件以“学号.html”形式命名）发给助教黄翰
（**huanghan@mail.bnu.edu.cn**）