吉林大学本科毕业设计（论文）文献综述

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学 院 | 计算机科学与技术学院 | | | | 专业 | 计算机科学与技术 |
| 学生姓名 | 孟政元 | | | | 学号 | **21140821** |
| 指导教师 | 王康平 | | | | 职称 | **讲师** |
| 合作导师 |  | | 职称 |  | 单位 | 计算智能 |
| 设计（论文）题目 | | 网易云音乐用户数据的分析和可视化技术研究与实现 | | | | |
| 文献综述（主要包括国内外现状、研究方向、进展情况、存在问题、参考文献等）（3000字以上）  *（说明：文献综述是通过系统地查阅与所选课题相关的国内外文献，进行搜集、整理、加工，从而撰写的综合性叙述和评价的文章。要体现“综合性”、“描述性”、“评价性”的特征。主体部分的结构包括该课题的“研究历史”的回顾，“研究现状”的对比，以及研究的“发展趋势”）*   1. **国内现状**   大数据时代下，信息技术突飞猛进，信息生产以几何级数增长，如何处理信息，更好地利用信息，成为各行各业所面临的问题。数据可视化，即通过对数据集的转化进行可视化设计，可以说是呈现数据的主要方式。常用的可视化工具主要分为信息图表类工具、基于时间顺序的时间线类工具和数据地图类工具。数据可视化在化繁为简，揭示新闻要点，梳理新闻进程；挖掘“数据关系”，揭示深层现象；符合受众习惯，及时转化受众反馈信息三方面具有重要意义。同时数据可视化在反映地区文化，展现信息随着时间的推移而产生的变化上也有一些作用。在数据可视化的操作要领中，一方面是在可视化设计中应该注意对数据的验证和数据的处理，另一方面在可视化的设计中，应该注意掌握可视化设计的几个要点，并不是复杂的设计就是好的设计，应该追求设计的简单大方，信息梳理准确，受众易获得信息。另外，相较国外的可视化实践，国内的可视化实践也在不断进行着，同时相对于互联网媒体而言传统媒体在可视化设计上优势不足，在这一问题上，传统媒体也应该思考如何突破重围。[1]   1. **研究方向**   目前数据可视化已经提出了许多方法，这些方法根据其 可视化的原理不同可以划分为基于几何的技术、面向像素技术、基于图标的技术、基于层次的技术、基于图像的技术和 分布式技术等等。[2]  用于创建和操作的可视化技术由数据集合生成 的图形描述 。有些可视化技术是针对某些特别的应 用开发的 ,而另一些技术具有普遍的适用性。这一部 分主要针对通用的可视化技术 。此外, 可视化技术涵 盖范围较广,这里只将可视化技术按一般可视化所必 需的过程划分为“数据预处理” 、“映射” 、“绘制”和“显 示”四步[3]  在可视化技术刚出现的时候,主要是用于数据可视化方面。现代的数 据可视化技术指的是运用计算机图形学和图像处理技术 ,将数据转换为图形 或图像在屏幕上显示 出来 ,并进行交互处理的理论、方法和技术。它涉及到计算机图形学、图像处理、计算机辅助设计、计算机视觉及人机交互技术等多个领域。科学家们不仅 需要通过图形图像来分析 由计算机算出的数据,而且盆要 了解在计算过程 中数据的变化。随着计算机技术 的发展,数据可视化概念也在不断地扩展 ,它不仅包括科学计算数据的可视化 ,而且包括工程数据的可视化。随粉网络技术和 电子商务 的发展,信息可视化便被提出 , 它 不仅能用圈像来显示多维的非空间数据 ,使用户加深对数据含义 的理解 ,而且用形象直观的图像来 指引检索过程 ,加快检索速度。随着可视化技术的不断发展 ,数据可视化与信息可视化的分界已越 来越不明显 。[4]   1. **进展情况**   人们可以在毫秒级的时间内从复杂的图像中抽取出信息 ,所以图像在传播信息的过程中比文字,声音更有效。因此可视 化技术在数据挖掘中具有重要 地位。可视化技术 按 目的可分 三类 : 探索型,验证型,表示型 。  探索型 : 人们事先没有关于数据的任何知识 ,而利用可视化技术分析数据 的结构,变化趋势,得到有关数据的假设 ;  验证型 : 人们事先有 关于数据 的假设 , 而利用可视化技术 验证或拒绝这些假设 ;  表示型: 选择有效的手段或技术表示数据 。[5]  基于 B/S 体系结构软件是各种应用软件发展的必然趋势, 在 B/S 体系结构下实现数据的可视化, 首先要考虑的就是哪种 图形文件格式是 Web 方式下的最佳选择。目前, 网络应用较多 的图形格式有 JPEG、PNG, 除此之外, 可伸缩矢量图形 SVG 在近几年也有了广泛的应用, 大有成为新一代 Web 图像标准之势。 下面就这几种图形格式作一下简单的比较  JPEG(Joint Picture Expert Group), 联合图像专家组,和任何 在 Web 上广泛使用的图像文件格式一样, JPEG 格式与平台无关, 它包含 24 位 RGB 颜色信息, 也就是说它们是由真彩色上万 种颜色空间的颜色组成, 可以获得较为丰富和精细的色彩变化。 在图像处理方面, JPEG 格式使用所谓的"有损"压缩技术, 即在压缩过程中将有某些颜色信息会被舍弃。对于 JPEG 图像, 在它们显示前要先进行解压, 这一过程要占用浏览器较多的时 间。在压缩过程中一旦损失了图形品质, 那么你将无法重新获 得, 图形品质的损失还可以累加, 即每次压缩或者解压图像,都会损失更多的图形信息。因此, 每次编辑图形或重新存储都会进 一步降低图形的品质。JPEG 格式特别不善于压缩上具有单一的 颜色区的图形, 如标志、印刷字体等。JPEG 的有损压缩会使单一 的颜色变得杂乱斑驳。JPEG 压缩也不善于压缩棱角边缘或印刷图形。  PNG(Portable Network Format), 可移植网络图像文件格式,是一种通用的全功能的图像文件格式。PNG 格式支持 8 位调色 板、16 位灰度级, 而且可达到 48 位 RGB 的支持, 而且 PNG 采用 无损压缩技术, 其图像文件的质量不会因为压缩而缺损, 这与用 牺牲一定图像质量为代价来减小文件大小的 JPEG 有损压缩方 式有根本的区别, 但是通常 PNG 文件要比 JPEG 文件要大一些。  SVG(Scalable Vector Graphics), 可伸缩的矢量图形, 它是由 W3C 制定的基于可扩展标记语言 (XML, Extensible Markup Language) 来描述二维适量图型的一个开放标准。SVG 严格遵从 XML 语法, 并用文本格式的描述性语言来描述图像内容, 是一 种和图像分辨率无关的适量图形格式, 因此图像文件的大小只 与图像本身的复杂程序有关, 而与图像的品质无关。 SVG 本身有具有许多前面两种图形格式无法比拟的优点, 如 SVG 图像文件可读, 易于修改和编辑; 可以嵌入 JavaScript 脚 本来控制 SVG 对象; 可以方便地建立文字索引, 从而实现基于 内容的图像搜索; 还可以用来动态生成图形, 实现交互式图形, 嵌入网页中, 并显示给终端用户。此外, SVG 特别适用于 Web 程 序, 可以由 Java Servlets 和 JSP 生成; 可以利用 JavaScript 脚本来 实现交互。 综上所述, 由于 SVG 图形具有可缩放、可搜索、更高的图形 品质、文件尺寸小、下载速度快、具有丰富的表现效果等优点, 使 之成为 Web 图形格式的首选。[6]   1. **存在问题**   在近十年时间里,数据采集、存储和数据分析技术飞速发展,大大降低了数据储存和处理的成本,使得一个大数据时代逐渐展现在我们的面前。曾经,枯燥的数据统计和分析只是统计学家、数据分析师和科研学者们的专利,而大数据革新性的将海量数据处理变为可能,并且大幅降低了成本,使得越来越多跨专业学科的人投入到大数据的开发应用中来。大数据能够帮助人们更明智的制定决策、更清晰的传达理念。面对大数据瑰丽且深奥的面貌,如何才能让大型数据集变得亲切和易于理解,可视化无疑是最有效的途径。对大数据背景下的数据可视化应用展开研究,将有助于我们发展和创新数据可视化技术。[7]  D3 是 Data-Driven Documents（数据驱动文档）的缩 写，是一种在数据文档 JavaScript 库基础上开发的可视化 数据处理技术，其功能的实现机制是将数据库中的数据与 HTML、SVG、CSS 结合起来，数据信息与这些规则的结 合能够让原本的二维数据，在立体展示空间内形成一个缜 密的相互连接的立体数据体系，全面的体现出数据信息之 间的相互关系，保证数据信息的具象可视化效果。在这一 过程中的数据主要来源于作者，文档代表基于 Web 的文 档或者网页文档，而 D3 在这一过程中扮演的角色是整个 驱动程序的推动者，在数据的可视化过程中将数据和文档 元素结合在一起。 在具体的数据可视化实践中 D3.js 采用的是链式语 法，能够很好的实现对数据库中线性数据的提取和应用。[8]   1. **参考文献**   [1] 李希娟. 大数据时代下的数据可视化研究[D]. 河北大学, 2014.  [2] 刘勘, 周晓峥, 周洞汝. 数据可视化的研究与发展[J]. 计算机工程, 2002, 28(8):1-2.  [3] 陈建军, 于志强, 朱昀. 数据可视化技术及其应用[J]. 红外与激光工程, 2001, 30(5):339-342.  [4] 石昊苏, 韩丽娜. 数据可视化技术及其应用展望[C]// 全国自动化新技术学术交流会会议论文集(一). 2005.  [5] 韩子良, 毕妤. 数据可视化在数据挖掘中的应用[J]. 计算机应用与软件, 2003, 20(11):71-73.  [6] 吴猛. 基于Web的数据可视化技术初探[J]. 福建电脑, 2007(12):58-59.  [7] 涂聪. 大数据时代背景下的数据可视化应用研究[J]. 电子制作, 2013, 47(5x):118-118.  [8] 赵聪. 可视化库D3.js的应用研究[J]. 信息技术与信息化, 2015(2):107-109. | | | | | | |

**注：**专业一定要填写全称。[专业名称：计算机科学与技术、计算机科学与技术（网络与信息安全）、物联网工程、理科试验班（唐敖庆计算机班）]