软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(5):1199−1211 [doi: 10.13328/j.cnki.jos.004959]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

一种模型驱动的可视化生成系统*

杜 一[1], 郭旦怀[1], 陈 昕[1], 任 磊[2], 戴国忠[3]

[1](中国科学院 计算机网络信息中心 科学数据中心,北京 100190)
[2](北京航空航天大学 自动化科学与电气工程学院,北京 100191)
[3](人机交互北京市重点实验室(中国科学院 软件研究所),北京 100190)
通讯作者: 杜一, E-mail: duyi@cnic.cn, http://www.cnic.cn

摘 要: 随着可视化与可视分析需求的增加,可视化开发所需的总成本不断提高.模型驱动的开发方法能够提高软件开发的效率,但在可视化开发中并没有发挥应有的作用.在此背景下,给出 DVDL(data visualization description language)——一种模块化、层次化描述的可视化描述语言,利用 DVDL 可以对组成可视化的各个部分进行不同抽象层次上的描述.在 DVDL 基础上,给出一种模型驱动的可视化生成系统 DVIZ(data visualization).该系统通过数据源选择、可视化配置、发布与分享这 3 个步骤以及所见即所得的配置,实现可视化的快速生成;系统支持多可视化图形之间的交互与关联配置;并且该系统基于 Web 开发,支持多种社交平台,易于分享与传播.

关键词: 可视化系统;用户界面描述语言;模型驱动的开发方法;最终用户编程
中图法分类号: TP391

中文引用格式: 杜一,郭旦怀,陈昕,任磊,戴国忠.一种模型驱动的可视化生成系统.软件学报,2016,27(5):1199−1211. http://www.jos.org.cn/1000-9825/4959.htm

英文引用格式: Du Y, Guo DH, Chen X, Ren L, Dai GZ. Model-Driven visualization generation system. Ruan Jian Xue Bao/Journal of Software, 2016,27(5):1199−1211 (in Chinese). http://www.jos.org.cn/1000-9825/4959.htm

Model-Driven Visualization Generation System

DU Yi[1], GUO Dan-Huai[1], CHEN Xin[1], REN Lei[2], DAI Guo-Zhong[3]

[1](Scientific Data Center, Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100190, China)
[2](School of Automation Science and Electrical Engineering, BeiHang University, Beijing 100191, China)
[3](Beijing Key Lab of Human-Computer Interaction (Institute of Software, The Chinese Academy of Sciences), Beijing 100190, China)

Abstract: While model-driven engineering (MDE) methodology has made significant improvements in terms of efficiency and effectiveness in many areas of software development, the same cannot be said of the development of data visualization systems. With this challenge in mind, this paper introduces DVDL (data visualization description language), a modular and hierarchical visualization description language that take advantage of the model-based design of MDE to describe visualization development at an abstract level. The paper also presents DVIZ (data visualization), a visualization system based on DVDL. With a growing popularity and demand for data visualization technology, a number of visualization tools have emerged in recent years, though few of them can be considered as adaptable and scalable as DVIZ. Key features in DVIZ include data source selection by user, property configuration of visual elements, and result publishing and sharing. The system also supports real-time result generation and multi-visual interaction. Lastly, since DVIZ is web-based, it supports result distribution across various social media.

Key words: visualization system; user interface description language; model-driven development methodology; end user programming

* 基金项目: 国家自然科学基金(61402435, 41371386, 61572057, 61303162); 中国科学院知识创新工程(CNIC_QN_1507)
Foundation item: National Natural Science Foundation of China (61402435, 41371386, 61572057, 61303162); Knowledge Innovation Program of the Chinese Academy of Sciences (CNIC_QN_1507)
收稿时间: 2015-07-31; 修改时间: 2015-09-19; 采用时间: 2015-11-10

© 中国科学院软件研究所 http://www.jos.org.cn

数据可视化的开发,需要实现包括数据转换、可视映射、视图转换[1]等多个步骤.随着人们通过数据可视化来探求数据背后的信息与知识的需求愈发迫切,可视化开发所需的总成本越来越多.模型驱动的开发方法通过对开发过程进行模型化的定义,能够很好地辅助开发,提高开发效率,并且已经在交互式应用程序的开发中得到了验证.Prefuse[2],D3.js[3]等可视化开发库及 Polaris[4],Lyra[5]等可视化开发工具,均在不同层次上体现了模型驱动的开发理念,提高了可视化开发的效率.但一方面,当前可视化开发中对模型的抽象、总结程度不够,大多数可视化开发工具对应的模型并不清晰,难以复用;另一方面,已有的应用于辅助可视化开发的模型对应的工具的易用性、可扩展性上存在不足.以上原因,导致模型驱动的开发方法在可视化开发领域中并没有发挥应有的作用.

本文面向可视化应用开发,基于模型驱动的理论,给出一种可视化应用模型 DVDL.该模型扩展 E-UIDL 用户界面描述语言,E-UIDL 通过模块化、层次化的设计方法将用户界面模型分成抽象功能界面描述模块、具体用户界面描述模块等 7 个模块.DVDL 在继承了相关模块的基础上,对具体可视化模块及具体可视化数据模块进行了定义.在 DVDL 基础上,给出可视化生成系统 DVIZ(data visualization),DVIZ 通过数据源配置、可视化配置、结果发布与分享这 3 个主要的功能模块,实现快速、灵活的可视化应用的生成.DVIZ 的主要特点包括:

(1) 通过 DVDL 指定的数据模型,实现了对通用数据服务的支持;

(2) 所见即所得的配置,能够实时观察配置后的可视化结果;

(3) 支持多可视化图形之间的交互与关联配置;

(4) 基于 Web 开发,支持多种社交平台,易于分享与传播.

本文的主要贡献在以下两个方面:(1) 基于可视化应用模型,给出一种可视化描述语言 DVDL,用来指导可视化应用的开发;(2) 给出基于 DVDL 的可视化生成系统 DVIZ,用来支持可视化的快速生成.

本文第 1 节从模型驱动的开发方法及可视化辅助开发方法两个方面对相关工作进行调研.第 2 节介绍 DVDL 模型.第 3 节给出基于 DVDL 的可视化生成系统 DVIZ 的系统设计与实现.在第 4 节中,通过应用实例对 DVDL 及 DVIZ 进行初步验证.最后,给出讨论及下一步工作.

1 相关工作

1.1 模型驱动的开发方法及描述语言

模型驱动的开发方法通过在软件开发过程中引入模型的概念,对软件开发过程中的各个部分进行抽象;以抽象后的模型为基础,通过各种辅助开发工具的支持,最终提高开发的效率.模型驱动的开发方法已在交互式应用程序的开发中得到验证[6−8].用户界面描述语言(user interface description language)以模型驱动的开发方法为基础,允许界面开发人员和设计人员使用一种规范的方式对用户界面进行描述,并利用各种生成工具生成用户界面或最终代码.XIML(extensible interface markup language)[9]定义了多种模型,并通过关系和属性使得各个独立的模型相互联系.UsiXML(user interface extensible markup language)[10]将界面抽象成多种模型,在该语言中,对具体界面模型中的界面组件的描述非常丰富.MARIA[6]以 TERESA 语言为基础,可以灵活地实现界面语言到相关实现的转换,并且对 Web 服务有较好的支持.除面向传统图形用户界面开发的描述语言外,还有一些针对非图形用户界面的描述语言.PUC[7]是一种针对电器设备控制界面描述的界面语言,利用这些语言,可以自动生成桌面、手持设备等多种设备上的图形用户界面或语音用户界面.它能够描述很多复杂的电器,并且可以根据设备的不同属性生成不同的控制界面.E-UIDL[11,12]借鉴优秀的用户界面描述语言设计上的特点,针对其在可扩展性、抽象能力、多设备支持上的不足,给出了新的用户界面描述语言及相关辅助工具.

在可视化开发中,van Wijk[13]针对可视化开发中的交互需求,设计出可以对可视化中的过渡及导航进行描述的语言,用以支持可视化中的交互.该语言侧重从算法层面首先对可视化中的部分交互过程建立数学模型,并给出相应描述.VisQL[14]从 Polaris[4]系统中演化出来,将数据的查询、分析及可视化进行整合,支持对多种不同可视化类型的描述,如表格、图形、地图、时序关系等.VisQL 不断演进,支持了 Tableau 系统的设计与实现.但该语言及其演化版本没有公开,无法了解其设计细节,对其可扩展性等特性无法评估.Heer 基于相关的可视化工具[2,3,15]及系统[5]的开发经验,结合声明式编程理念,给出了可视化中的声明式语言设计[16,17],并给出了一种用于

© 中国科学院软件研究所 http://www.jos.org.cn

设计可视化的声明式语言 Vega.该语言对可视化进行了充分的抽象,将一个可视化分成数据(data)、数据转换(data transforms)、刻度(scales)、导航(guides)、标记(marks)等主要组成部分,并给出了 Vega 的 JSON 结构.Vega 及其相关的系统及工具,是可视化应用开发中的有益尝试.但 Vega 设计时主要考虑了对单个可视化本身的抽象,对多可视化之间的交互以及基于 Vega 的扩展方面,依然存在一些挑战.

1.2 可视化辅助开发

为了辅助可视化的开发,出现了各种不同的工具及系统.这些工具及系统可以划分为 3 个不同的层次:第 1 层次为基本图形库,这些图形库包括了对图形渲染的支持,如 OpenGL,Processing,Java2D 等;第 2 层为可视化开发工具,可视化开发工具在基本图形库的基础上实现了诸如不同的可视化方法、布局类型、颜色映射等功能,在一定程度上简化了可视化开发人员开发可视化的复杂度,这类开发工具包括 Prefuse[2],D3.js[3],Protovis[15],TimeBench[18],Flare,VTK,ECharts 等;第 3 层为面向最终用户编程的可视化开发系统,这类系统面向可视化设计人员或最终用户,通过直观的操作,实现可视化的生成.基本图形库、可视化开发工具及可视化开发系统,随着层次的推进,其开发复杂程度逐渐降低,但由于过程中对功能的抽象及封装,也导致其灵活性的下降.本文实现的 DVIZ 处于第 3 层,在设计与实现 DVIZ 时,对可视化开发系统进行了详细调研.

SAGE[19]是比较经典的可视化设计系统,系统可以根据数据的特性,在使用者指定或部分指定相关约束的情况下,自动生成可视化,极大地降低了可视化设计的工作量.但是,SAGE 基于数据本身的特性进行可视化的生成,并且使用图元、坐标轴等作为系统可视化生成的原语,对于缺乏设计经验的使用者来说,易用性并不理想.FLINAView[20],iVisDesigner[21],Lyra[5]与 SAGE 在设计思路上部分类似,均对可视化进行了抽象,基于 Bertin[22]的图形设计准则,对将可视化分成了图形、标记、颜色等不同的部分,并分别设计.FLINAView 实现了对基于坐标轴的可视化的设计,能够快速设计出风格各异的平行坐标.该系统交互流畅,对多维数据可视化设计系统有很高的借鉴意义.iVisDesigner 是一款基于 Web 的可视化设计工具,基于其开发的可视化极具表现力.FLINAView 及 iVisDesigner 设计的可视化作品均支持经典的交互操作,如格式刷、拉锁等.Lyra 在功能上与 iVisDesigner 类似,但该系统基于 Vega,有完善的描述语言的支持,具有很好的可扩展性.

Many Eyes[23]从可视化方法出发,允许使用者在上传数据后,首先指定可视化方法,并在此基础上进行各种配置,最终生成交互式可视化结果.该系统在设计时给出了相应的数据模型,理论上可以方便地进行新的可视化方法的扩展,但系统对可视化模型没有指定,并且对交互及多个可视化之间的关联表现不好.Polaris[4]在图形层面给出了描述,通过对散点图的支持,实现了对多维数据的分析及可视化.基于 Polaris 的 Tableau 系统,在商业上取得了极大的成功.PanoramicData[24]以数据为驱动,将纸笔隐喻引入可视化分析与设计中,通过手势交互,快速实现可视化的分析,并且对关联分析有良好的支持,但该系统仅支持基本的可视化方法,且系统没有模型支持,在可扩展性上并不明朗.

从基本的坐标轴、图元进行设计的可视化设计与生成系统,在灵活性上比基于可视化方法的系统具有优势,但在功能可用性(affordance)上不及基于可视化方法的系统,因此更适合具有设计经验的人.本文给出的 DVIZ 系统属于后者.上述可视化设计系统,除 Many Eyes 外,均不支持结果的分享,或分享仅限于对可视化结果导出的图片的分享,不利于可视化的传播.

相对于已有的基于可视化方法的系统,本文实现的可视化设计及生成系统 DVIZ 有了模型的支持,在可扩展性上具有优势;并且在可视化结果的交互、多个可视化之间的关联上有良好的支持.此外,DVIZ 还能支持多种社交平台,便于将交互式可视化结果分享.

2 DVDL

E-UIDL[12]是一种可扩展的用户界面描述语言,使用模块化的描述方式,包括抽象界面模型(AFUI)、具体界面模型(CUI)、抽象数据模型(ADATAMODEL)、具体数据模型(CDATAMODEL)、用户模型(UM2)、映射模型(MAPPING)及资源模型(RESOURCE),支持对界面的多层次描述,且具有无关性及良好的可扩展性、可复用性等特点.基于该描述语言,可以实现辅助笔式用户界面的开发、多设备界面自动生成[11]、自适应用户界面设计

© 中国科学院软件研究所 http://www.jos.org.cn

等.本文借鉴E-UIDL成功的设计经验,结合可视化的特点对E-UIDL进行扩展,给出可视化设计对应的描述语言DVDL.

2.1 设计特性

DVDL 在设计时继承了 E-UIDL 的一些特性.首先,DVDL 使用模块化的描述方式进行描述,对可视化的描述包括对数据源、整体布局、可视化方法、详细配置、关联等多个不同部分的描述,通过模块化的描述方法,不仅能够增加 DVDL 本身的可读性,而且能够增加对组成可视化的各个模块的复用,而复用的能力则取决于 DVDL 的模块化的划分粒度;其次,DVDL 支持多层次的描述,通过在不同抽象层次上的描述,可以实现基于 DVDL 的可视化推荐,从而提高可视化的开发效率;此外,DVDL 具有良好的可扩展性.新的可视化及交互技术不断出现,通过良好的可扩展性,使得 DVDL 不断支持新的可视化及交互技术.

2.2 组成模块

基于模块化、层次化描述及良好可扩展性的设计原则,我们给出 DVDL 的描述.图 1 给出 DVDL 主要部分的 XML Schema 结构图,图中每个矩形框表示一个节点元素,矩形框跟随的加号和减号分别表示该元素是否展开,矩形框下方的数字代表节点允许出现的次数.我们分别使用内容为 S,C,A 的矩形框表示 XML Schema 描述时的 Sequence,Choice 及 All 这 3 种模型,其中,Sequence 模型表示所有元素依次出现,Choice 模型表示所有元素选择一项出现,All 模型表示所有元素全部出现.

DVDL 是描述语言的根结构,它包括整个可视化描述文件的基本参数,如可视化标示符、名称、版本信息等.如图 1 所示,DVDL 主要由 5 个子模块组成,包括抽象功能界面描述模块(AFUI)、具体可视化模块(CV)、抽象数据描述模块(ADATAMODEL)、可视化数据模块(CVDATAMODEL)以及映射模块(MAPPING).DVDL 与每个子模块均为一对多的关系,即,DVDL 可以包括 0 个或多个子模块的实体.如在本文给出的 DVIZ 系统中,每一个界面描述文件各包括了一个或多个具体可视化模块、一个或多个可视化数据模块以及一个映射模块.

由于 E-UIDL 在描述用户界面时的合理抽象以及可视化开发在某些程度上与用户界面开发的类似,DVDL 在设计时继承了 E-UIDL 对 AFUI,ADATAMODEL 及 MAPPING 的描述.关于 E-UIDL 对 AFUI,ADATAMODEL 及 MAPPING 的描述,可以参考文献[12].由于在多数可视化系统描述中对用户模型及资源模型的使用较少,因此在 DVDL 的描述中,去掉用户模型及资源模型.

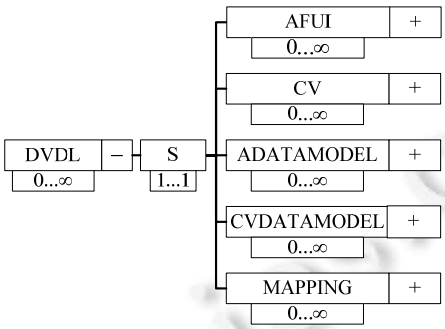


Fig.1 XML Schema of DVDL

图 1 DVDL 的 XML Schema

具体可视化模块是对可视化的具体描述.具体可视化模块与抽象用户界面描述模块是一种映射关系,一种抽象用户界面描述模块可有多种具体可视化模块的具体实现.使用两个单独的模块来描述可视化,可以保证可视化在一定抽象层次上的可复用性.如图 2 所示,具体可视化模块各个元素之间相对独立,便于添加新的可视化图形的支持.目前,该模块中定义的可视化图形包括折线图(LineChart)、柱状图(BarChart)、散点图(ScatterPlot)、饼图(PieChart)、雷达图(RadarChart)、和弦图(ChordChart)、*K* 线图(KLineChart)、力导向图(ForceGraph)、区域地图(AreaMap)、点地图(PointMap)、线地图(LineMap)及自定义图形(Customized).图 3 为对散点图的描述,

© 中国科学院软件研究所 http://www.jos.org.cn

通过一定程度的抽象,对一些可视化图形中公共的部分进行了提取,其中包括标题(title)、图例(legend)、范围选择框(range)等.除公共部分外,散点图还包括对轴(axis)及点(point)的映射的配置.其中,轴的配置包括 X 轴(XAxis)、Y 轴(YAxis)对应的维度,点的配置包括点的尺寸(size)、颜色(color)、形状(shape)分别对应的维度.通过该描述,可以指定散点图各个图形属性对应的数据以及图形与数据之间的映射关系.

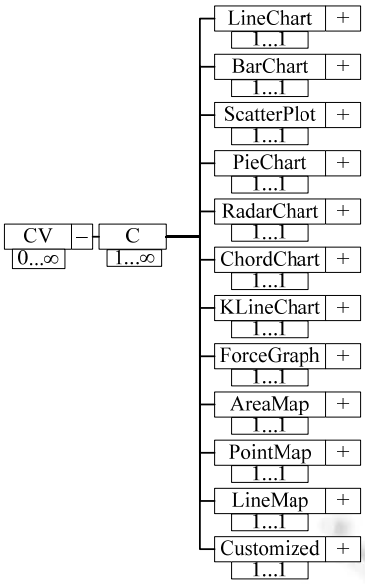


Fig.2 XML Schema of CV module

图 2 CV 模块的 XML Schema

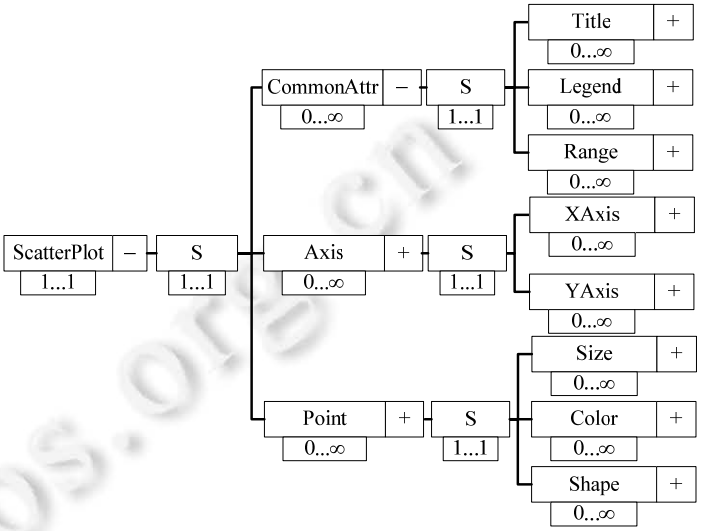
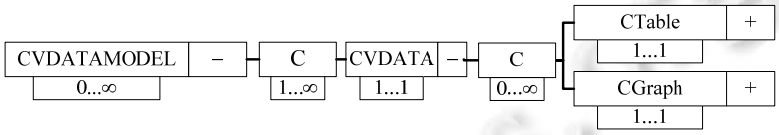


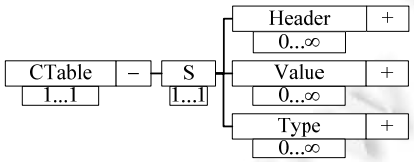
Fig.3 XML Schema of ScatterPlot in CV module

图 3 CV 模块中 ScatterPlot 的 XML Schema

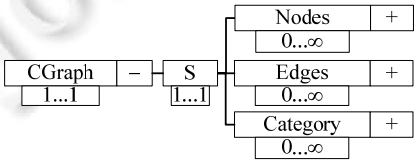
可视化数据模块是 DVDL 的具体数据描述模块,它可以对用于进行可视化的数据进行详细具体的描述.在当前版本的描述中,为了更加灵活地支持多种数据源,我们定义了两类数据,分别是表格结构和网络结构,且数据类型可根据需求进行扩展.图 4 分别呈现了 CVDATAMODEL 的 XML Schema 的描述,其中,表格结构包括对表头、数值及类型的定义,网络结构包括了对节点、边及类型的定义,两种结构中的数据类型包括数值型、类别型、日期、经纬度.由于表格结构与网络结构的互通性[2],在 DVIZ 中实现了网络结构到表格结构的自动转换.



(a) CVDATAMODEL 模块的 XML Schema



(b) 表格结构的 XML Schema



(c) 网络结构的 XML Schema

Fig.4 XML Schema of CVDATAMODEL, table and graph

图 4 CVDATAMODEL 模块及表格、网络结构的 XML Schema

© 中国科学院软件研究所 http://www.jos.org.cn

3 系统设计与实现

3.1 系统框架

基于 DVDL 的描述,我们给出可视化生成系统 DVIZ.该系统使用了 DVDL 中的具体可视化模块、可视化数据模块以及映射模块,其系统框架如图 5 所示.系统以 DVDL 为基础,包括数据处理引擎、可视化引擎、结果生成引擎这 3 个主要模块.

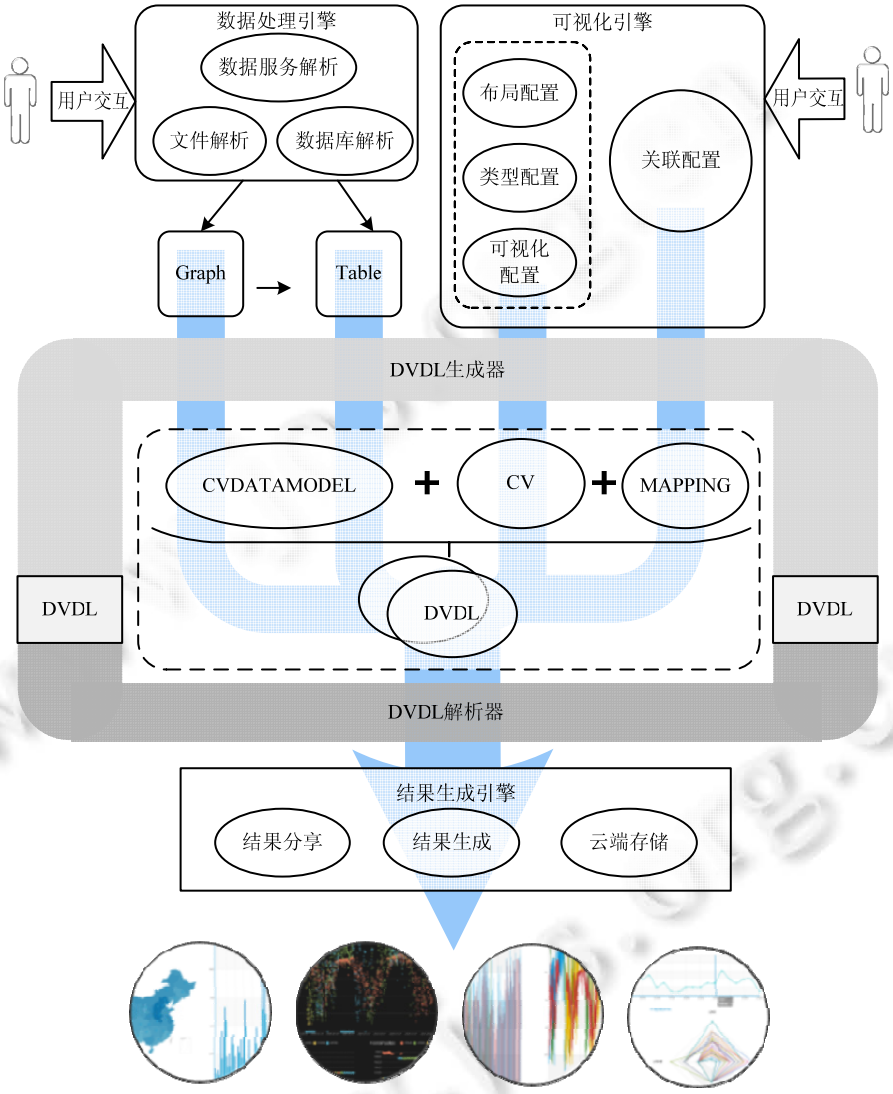


Fig.5 Framework of DVIZ

图 5 DVIZ 体系框架

数据处理引擎接收用户指定的数据源,并根据指定的规则对数据源进行处理,在数据处理时,根据输入的数据的情况,可以将数据源转换成表格结构或网络结构数据.由于表格结构与网络结构的互通性,网络结构数据可以根据需求自动转换成表格数据.数据处理引擎处理后的结果,通过 DVDL 生成器转换成 CDATAMODEL 模型存储,用于支持后续模块的操作.数据处理引擎包括文件解析、数据库解析、数据服务解析这 3 个子模块,可以

© 中国科学院软件研究所 http://www.jos.org.cn

对 3 类不同的数据源进行解析.其中,文件解析模块实现了对.CSV,.TXT,.XLS,.XLSX 这 4 种表格结构文件的解析以及对.GML,.GEXF 这两种网络结构文件的解析;数据库解析模块实现了对 Mysql,Oracle,SQLServer 这 3 种关系型数据库的解析以及对 MongoDB 这种 NoSQL 数据库的解析;数据服务解析模块支持静态服务以及动态数据服务,可以将第三方数据服务作为数据源进行解析.

可视化引擎是 DVIZ 的核心模块,它接收数据处理引擎生成的 DVDL 描述的 CDATAMODEL 部分,并对可视化进行各种配置,最终生成 DVDL 描述,并支持后续的存储与分享.该模块包括布局配置、类型配置、可视化配置、关联配置,其中,布局配置用以指定可视化结果整体的布局,如可视化图形或控件的数目;类型配置用以指定各个可视化图形或控件的类型;可视化配置用以详细配置各个可视化图形的具体参数,如坐标轴对应的数据、可视化图形元素的颜色、尺寸等属性;关联配置用以配置多个不同的可视化图形之间,或可视化图形与控件的关联.在可视化引擎的各个模块中,布局配置、类型配置及可视化配置的结果,通过 DVDL 生成器处理后被写入 CV 模型;关联配置的结果通过 DVDL 生成器被写入 MAPPING 模型存储.可视化引擎处理后的 DVDL 模型作为结果生成引擎的输入,生成可视化并分享.

结果生成引擎读取 DVDL 模型,并最终生成可视化并分享.该部分包括结果生成、结果分享及云端存储这 3 个子模块.结果生成模块获得数据处理及可视化引擎生成的 DVDL 描述,通过 DVDL 解析器,渲染并生成可视化结果;可视化结果通过云端存储模块进行存储,并给出可以通过 Web 访问的结果页面的链接;结果分享模块将可视化结果进行包装后,分享到微博、朋友圈、腾讯 QQ 等社交网络中.

3.2 系统界面

图 6 给出了 DVIZ 的系统界面,如图所示,系统使用包括数据源选择、可视化配置、发布与分享这 3 个步骤.图 6(a1)为选择数据源类型,图 6(a2)为数据源加载与预览,图 6(b1)为可视化布局选择,图 6(b2)为可视化图形配置,图 6(c1)为可视化结果预览,图 6(c2)为可视化结果分享.

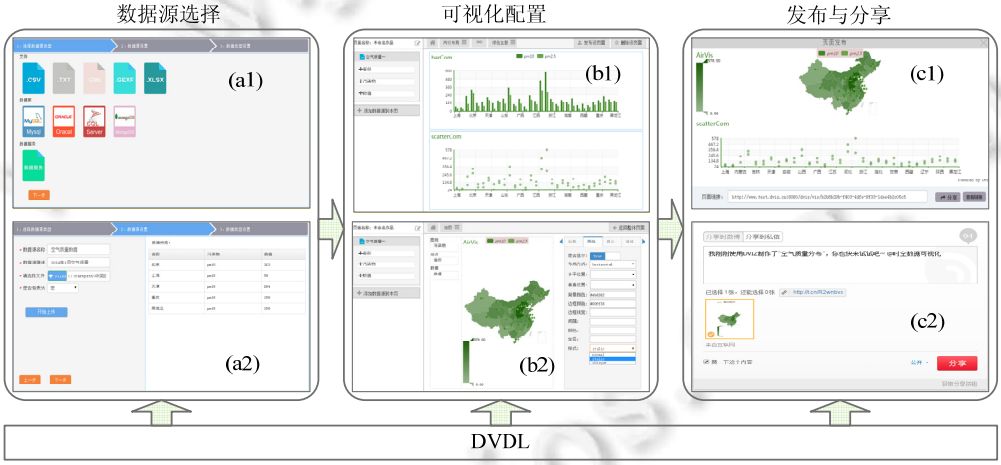


Fig.6 System interface of DVIZ

图 6 DVIZ 系统界面

3.2.1 数据加载

DVIZ 引入可视化项目的概念,每个可视化项目包括数据源及可视化页面两个部分.其中,数据源用来存储与展现已加载的数据源,可视化页面用来展现设计完成的可视化结果.可视化项目的引入可以提高对可视化项目中的数据、页面的管理效率.

使用者在新建可视化项目后,可以加载已有数据,DVIZ 在 DVDL 支持下,提供了对 5 种文件、3 种关系型数据库、一种非关系型数据库以及两种数据服务的支持.加载数据功能包括 3 个步骤,如图 6 所示.首先选择数据

© 中国科学院软件研究所 http://www.jos.org.cn

源类型(如图 6(a1)所示),然后制定数据的细节并加载(如图 6(a2)所示)、预览数据(如图 6(a2)所示),最后对数据类型进行调整.在 3 个步骤中,每一步都会对 CDATAMODEL 进行更新.以某个 CSV 文件为例,在该步骤后,更新的 CDATAMODEL 如图 7 所示.在 CDATAMODEL 中,定义了唯一标示符、数据描述等通用属性以及数据的类型,数据包含的列名、数据的详细内容以及各列数据类型等.

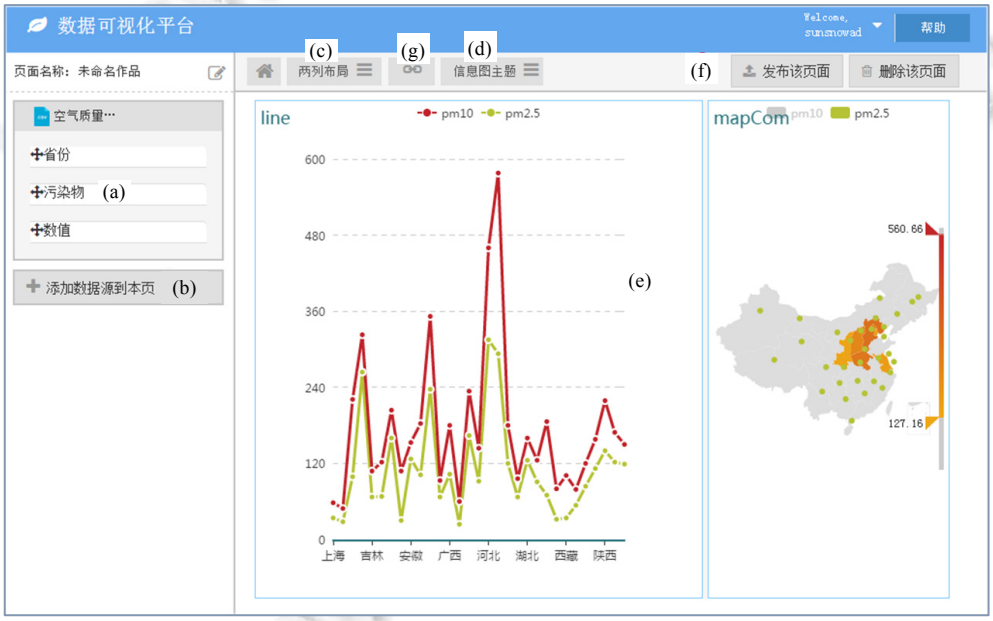
```
{
    id: “cdata1dee4b2c05c5”,
    desc: “cdatasample”,
    type: “ctable”,
    header: [“Car”,“MPG”,“Cylinders”,“Displacement”,“Horsepower”,“Weight”,“Acceleration”,“Model”,“Origin”],
    value: [
      [“Chevrolet Chevelle Malibu”,18,8,307,130,3504,12,70,“US”],
      [“Buick Skylark”,320,15,8,350,165,3693,11.5,70,“US”],
      ...
      ]
    type: [“nominal”,“ordinal”,“ordinal”,“ordinal”,“ordinal”,“ordinal”,“ordinal”,“ordinal”,“nominal”]}}
```

Fig.7 CDATAMODEL instance in JSON format

图 7 JSON 格式的 CDATAMODEL 实例

3.2.2 可视化配置

可视化配置可在数据加载完成后,或在可视化项目中选择数据后进行操作,进入可视化配置界面.该界面包括数据源的选择与重置、布局选择、关联配置、主题选择及主显示区域几个部分.如图 8 所示,数据列呈现一方面可以帮助直观了解当前数据情况,另一方面可以在单个可视化配置中通过“drag-and-drop”的方式实现所见即所得的可视化配置.数据源选择与重置部分可以显示已选择的数据源的数据情况,系统支持在同一个可视化结果中使用多个不同的数据源,可以根据需要替换数据源.DVIZ 提供了 4 种默认的布局供使用者选择,它允许使用者选择任一可视化结果的布局,并且支持在基础布局的基础上进行所见即所得的调整.布局选择与调整后,可以对单个布局区域进行可视化图形的设置.



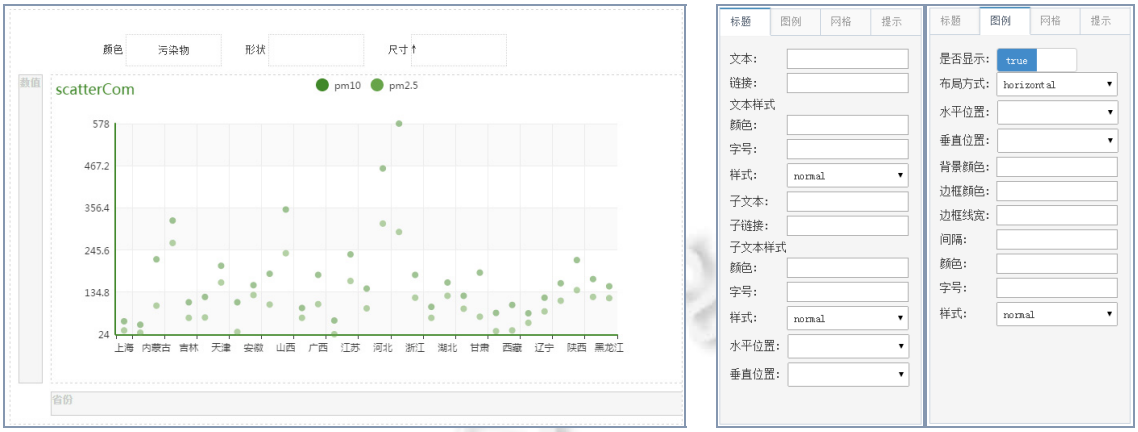
(a) 数据列呈现; (b) 更换或重新加载数据源; (c) 更换可视化布局; (d) 更换主题
(e) 配置结果显示区域; (f) 预览与发布; (g) 关联配置

Fig.8 Visual configuration interface

图 8 可视化配置界面

© 中国科学院软件研究所 http://www.jos.org.cn

对于不同的可视化图形,支持不同的配置项;对于不同的可视化图形的配置,通过点击主显示区域的图形编辑按钮进入.图 9 给出了散点图的配置界面.该功能包括两个部分:第 1 部分(如图 9(a)所示)是可以通过将任意数据列拖拽至指定区域,实现数据列与散点图图形属性的映射,包括 X 轴、Y 轴、点的颜色、尺寸、形状等;第 2 部分(如图 9(b)所示)为详细配置,可以为散点图进行更加详细的配置,如,可以对散点图的标题、图例、网格显示、高亮等进行配置.



(a) 散点图预览及属性映射界面 (b) 散点图的标题及图例配置界面

Fig.9 Configuration interface for a scatter plot

图 9 散点图配置界面

如果一个可视化结果中包含多个可视化图形,则需要对多个图形之间的关联进行配置.关联配置允许在结果包含多个可视化图形时,将两个图形对应的数据中的某个维度的数据进行关联.当在一个图形上交互,如鼠标点击图元、图例或使用格式刷选择区域范围时,交互的数据发生变化,另一个图形可以根据数据的变化而重新绘制.为降低开发可视化的成本,DVIZ 提供了主题选择的功能,通过选择不同的主题风格,可以保证缺乏设计经验的使用者也能设计出风格统一且和谐的可视化结果.主显示区域部分用以显示已经配置完毕的布局及可视化结果.

3.2.3 结果发布

经过可视化配置后,在主显示区域部分可以对可视化配置的结果进行预览,并将理想的可视化结果进行发布,如图 6(c1)所示.发布的可视化结果将以唯一的 URL 对外提供服务,使用者可以将可视化结果嵌入到个人的网页、各种演示文稿以及论文中.为便于分享,发布后的可视化结果可以通过各种社交网络进行分享,如新浪微博、朋友圈等,如图 6(c2)所示.

3.3 系统实现

由于 JSON 本身在网络数据交换时的优势,在 DVIZ 实现时,使用 JSON 对 DVDL 进行描述,并将其作为模型实例贯穿 DVIZ 开发的始终.系统使用 J2EE 架构进行开发,实现后台数据处理、模型解析的工作.为了灵活地支持新的数据类型及可视化图形,使用了面向对象的设计思想,其中,数据处理与解析的类图如图 10 所示.在整个数据相关的类中,DataParser 为根接口,TableDataParser 及 GraphDataParser 为实现了 DataParser 的两个抽象类,各个不同数据类型的解析与加载,通过继承相应的 TableDataParser 及 GraphDataParser 来实现.为了实现 DVDL 模型到可视化结果的呈现,在结果生成引擎中使用了模板引擎 FreeMarker,根据布局、可视化图形、关联等可视化结果的不同部分,实现了层次化的模板.

系统前端使用 HTML5 和 JavaScript 进行界面及可视化功能的实现.开发中利用了 Jquery,D3.js,ECharts 等第三方 JavaScript 库来实现文档管理、可视化渲染等功能.

© 中国科学院软件研究所 http://www.jos.org.cn

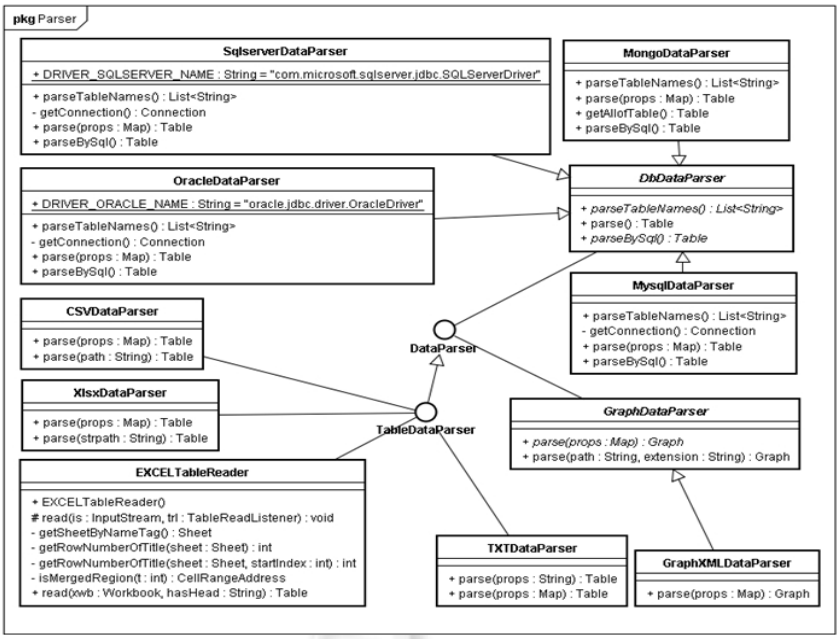


Fig.10 Class diagram of DataParser class

图 10 数据解析类图

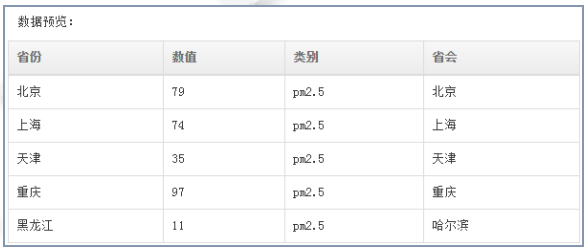
4 应用实例

4.1 多种数据类型

DVIZ 支持多种不同的数据类型,对数据服务的支持有两种不同的类型,分别是静态服务与动态服务.其中,静态服务指数据服务的参数不随时间变化,动态服务指数据服务的一个或多个参数随时间变化而变化.对静态服务的配置界面如图 11 所示.在本实例中,我们给出如何将数据服务作为数据源,设计并生成可视化.本实例中使用两个不同的数据服务 service1 及 service2,其中,service1 提供全国当前的空气质量情况,包括一个参数为 type,可以选择不同的空气质量指标.在服务连接后,在数据预览窗口会显示数据预览的效果,如图 12 所示.随后可以进行可视化的设计与实现.

Fig.11 Static data service configuration interface

图 11 静态服务配置界面



数据预览:

省份	数值	类别	省会
北京	79	pm2.5	北京
上海	74	pm2.5	上海
天津	35	pm2.5	天津
重庆	97	pm2.5	重庆
黑龙江	11	pm2.5	哈尔滨

Fig.12 Data preview interface

图 12 数据预览界面

对动态服务的配置界面如图 13 所示.该实例显示了该数据 2015 年 1 月 1 日~2015 年 1 月 31 日全国的 PM2.5

© 中国科学院软件研究所 http://www.jos.org.cn

情况.在动态服务连接之后,在数据预览窗口会显示参数在起始值对应的服务返回的数据的预览.之后,可以进行可视化的设计与实现.图 14 给出了使用动态服务配置生成的可视化结果.该结果随时间变化而动态变化.

Fig.13 Dynamic data service configuration interface

图 13 动态数据服务配置界面

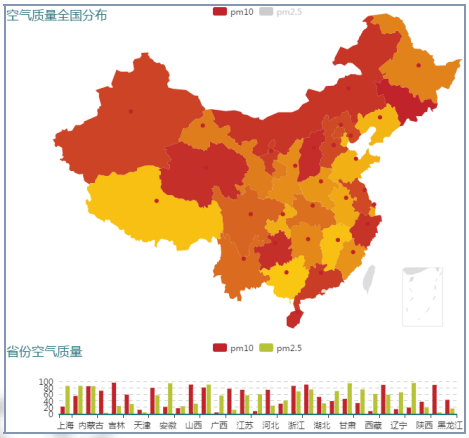


Fig.14 Visualization result preview

图 14 可视化结果预览

4.2 多数据关联

DVIZ 提供对多数据源的支持以及对多可视化图形之间关联的支持.在可视化设计的过程中,可以随时选择新的数据源进行加载.在本例中,首先基于空气质量数据设计两个独立的可视化图形:第 1 个图形使用地图呈现不同的省份的空气质量情况,第 2 个图形使用柱状图呈现某一类空气质量指标的数值情况.为实现在地图中选择特定的空气质量指标后,柱状图能够实时呈现该指标的数值情况,进行关联配置,图 15 显示了该实例关联的配置界面,首先,通过拖拽操作选择进行关联的图形,然后对关联的数据维度、关联动作进行制定,在关联配置完毕后,可以对结果进行预览及发布.图 16 显示了发布后的结果,在地图中选中 PM2.5 后,右侧呈现 PM2.5 指标的数值情况.

Fig.15 Association setting interface

图 15 关联配置界面

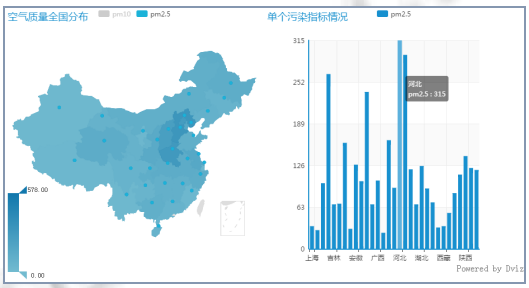


Fig.16 Visualization result preview with data association of legend

图 16 图例数据关联可视化结果预览

5 讨 论

本文首先给出了一种扩展了 E-UIDL 的可视化描述语言 DVDL,在此基础上,给出了基于 DVDL 的可视化设计与生成系统 DVIZ.该系统利用 DVDL 中的具体可视化模块、可视化数据模块以及映射模块,实现了可视化的快速生成.系统通过所见即所得的配置,实现可视化结果的实时观察;系统支持数据服务,使得系统可以设计随时间、随参数变化的可视化图形;系统支持多可视化图形之间的交互配置,可以实现多个可视化图形之间的

© 中国科学院软件研究所 http://www.jos.org.cn

关联;最后,DVIZ 基于 Web 开发,能够支持多种社交平台,易于分享与传播.

相对于其他可视化生成系统,DVIZ 最大的特点是基于一种可扩展的描述语言 DVDL.将 DVDL 作为核心,可以基于此标准实现不同的可视化生成系统,这些系统面向具有不同背景的角色,如可视化设计师、统计学家等.而且,各个系统之间可以进行相互转换,例如,系统 *A* 与 *B* 均基于 DVDL,分别面向统计学家及可视化设计师.通过将统计学家利用系统 *A* 设计的可视化结果导入到系统 *B* 中,可以允许可视化设计师在已有的可视化基础上进一步进行细节的设计,从而降低统计学家与可视化设计师在协作生成可视化时的沟通成本,提高可视化设计师的工作效率[25].DVDL 本身模块化的设计方法,也为实现多系统之间的转换提供了可能.

虽然 E-UIDL 已经证明了模块化描述用户界面可以实现新的界面元素、交互组件的扩展,但可视化的开发有其本身的特点,如何进一步将 DVDL 的其他模块加以应用以实现可视化推荐等功能,将是下一步工作的主要内容.

相对于其他可视化生成系统,DVIZ 还存在一些不足:(1) 在对数据的操作上,DVIZ 不支持在配置界面对数据进行操作,例如,将已有数据的各种统计指标,如求和、平均值等在所见即所得的配置中体现;(2) 在图形支持的种类上,目前 DVIZ 只支持 9 种不同的可视化图形,在下一步工作中将添加对平行坐标等可视化图形的支持.

致谢 在此,我们向对本文的工作给予支持和建议的同行表示感谢.

References:

[1] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: Proc. of the IEEE Symp. on Visual Languages. IEEE, 1996. 336−343. [doi: 10.1109/VL.1996.545307]

[2] Heer J, Card SK, Landay JA. Prefuse: A toolkit for interactive information visualization. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. ACM Press, 2005. 421−430. [doi: 10.1145/1054972.1055031]

[3] Bostock M, Ogievetsky V, Heer J. Data-Driven documents. IEEE Trans. on Visualization and Computer Graphics, 2011,17(12): 2301−2309. [doi: 10.1109/TVCG.2011.185]

[4] Stolte C, Tang D, Hanrahan P. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. IEEE Trans. on Visualization and Computer Graphics, 2002,8(1):52−65. [doi: 10.1109/2945.981851]

[5] Satyanarayan A, Heer J. Lyra: An interactive visualization design environment. Computer Graphics Forum, 2014,33(3):351−360. [doi: 10.1111/cgf.12391]

[6] Paterno F, Santoro C, Spano LD. MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM Trans. on Computer-Human Interaction, 2009,16(4):Article 19. [doi: 10.1145/1614390.1614394]

[7] Nichols J, Myers BA. Creating a lightweight user interface description language: An overview and analysis of the personal universal controller project. ACM Trans. on Computer-Human Interaction, 2009,16(4):Article 17. [doi: 10.1145/1614390.1614392]

[8] Navarre D, Palanque P, Ladry JF, Barboni E. ICOs: A model-based user interface description technique dedicated to interactive systems addressing usability, reliability and scalability. ACM Trans. on Computer-Human Interaction, 2009,16(4):Article 18. [doi: 10.1145/1614390.1614393]

[9] Eisenstein J, Vanderdonckt J, Puerta A. Applying model-based techniques to the development of UIs for mobile computers. In: Proc. of the 6th Int'l Conf. on Intelligent User Interfaces. ACM Press, 2001. 69−76. [doi: 10.1145/359784.360122]

[10] Limbourg Q, Vanderdonckt J, Michotte B, Bouillon L, López-Jaquero V. USIXML: A language supporting multi-path development of user interfaces. EHCI/DS-VIS, 2004,3425:200−220. [doi: 10.1007/11431879_12]

[11] Du Y, Tian F, Ma CX, Dai GZ, Wang HA. A mobile user interface generation framework based on multi-scale description. Chinese Journal of Computers, 2013,36(11):2179−2190 (in Chinese with English abstract).

[12] Du Y, Deng CZ, Tian F, Ren L, Dai GZ. Extensible user interface description language. Ruan Jian Xue Bao/Journal of Software, 2013,24(5):1127−1142 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4321.htm [doi: 10.3724/SP.J.1001.2013.04321]

[13] Van Wijk JJ, Nuij WAA. A model for smooth viewing and navigation of large 2D information spaces. IEEE Trans. on Visualization and Computer Graphics, 2004,10(4):447−458. [doi: 10.1109/TVCG.2004.1]

[14] Hanrahan P. Vizql: A language for query, analysis and visualization. In: Proc. of the 2006 ACM SIGMOD Int'l Conf. on Management of Data. ACM Press, 2006. 721−721. [doi: 10.1145/1142473.1142560]

© 中国科学院软件研究所 http://www.jos.org.cn

[15] Bostock M, Heer J. Protovis: A graphical toolkit for visualization. IEEE Trans. on Visualization and Computer Graphics, 2009, 15(6):1121−1128. [doi: 10.1109/TVCG.2009.174]

[16] Satyanarayan A, Wongsuphasawat K, Heer J. Declarative interaction design for data visualization. In: Proc. of the 27th Annual ACM Symp. on User Interface Software and Technology. ACM Press, 2014. 669−678. [doi: 10.1145/2642918.2647360]

[17] Heer J, Bostock M. Declarative language design for interactive visualization. IEEE Trans. on Visualization and Computer Graphics, 2010,16(6):1149−1156. [doi: 10.1109/TVCG.2010.144]

[18] Rind A, Lammarsch T, Aigner W, Aigner W, Alsallakh B, Miksch S. TimeBench: A data model and software library for visual analytics of time-oriented data. IEEE Trans. on Visualization and Computer Graphics, 2013,19(12):2247−2256. [doi: 10.1109/TVCG.2013.206]

[19] Roth SF, Kolojejchick J, Mattis J, Goldstein J. Interactive graphic design using automatic presentation knowledge. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. ACM Press, 1994. 112−117. [doi: 10.1145/191666.191719]

[20] Claessen JHT, Van Wijk JJ. Flexible linked axes for multivariate data visualization. IEEE Trans. on Visualization and Computer Graphics, 2011,17(12):2310−2316. [doi: 10.1109/TVCG.2011.201]

[21] Ren D, Hollerer T, Yuan X. iVisDesigner: Expressive interactive design of information visualizations. IEEE Trans. on Visualization and Computer Graphics, 2014,20(12):2092−2101. [doi: 10.1109/TVCG.2014.2346291]

[22] Bertin J. Semiology of Graphics: Diagrams, Networks, Maps. Esri Press, 1983.

[23] Viegas FB, Wattenberg M, Van Ham F, Kriss J, McKeon M. Manyeyes: A site for visualization at internet scale. IEEE Trans. on Visualization and Computer Graphics, 2007,13(6):1121−1128. [doi: 10.1109/TVCG.2007.70577]

[24] Zgraggen E, Zeleznik R, Drucker SM. PanoramicData: Data analysis through pen & touch. IEEE Trans. on Visualization and Computer Graphics, 2014,20(12):2112−2121. [doi: 10.1109/TVCG.2014.2346293]

[25] Du Y, Tian F, Dai GZ. Development approach based on extensible user interface description language. Ruan Jian Xue Bao/Journal of Software, 2015,26(7):1772−1784 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4584.htm [doi: 10.13328/j.cnki.jos.004584]

附中文参考文献:

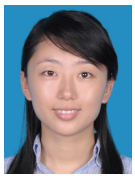
[11] 杜一,田丰,马翠霞,戴国忠,王宏安.基于多尺度描述方法的移动用户界面生成框架.计算机学报,2013,36(11):2179−2190.

[12] 杜一,邓昌智,田丰,任磊,戴国忠.一种可扩展的用户界面描述语言.软件学报,2013,24(5):1127−1142. http://www.jos.org.cn/1000-9825/4321.htm [doi: 10.3724/SP.J.1001.2013.04321]

[25] 杜一,田丰,戴国忠.E-UIDL 用户界面描述语言下的开发方法.软件学报,2015,26(7):1772−1784. http://www.jos.org.cn/1000-9825/4584.htm [doi: 10.13328/j.cnki.jos.004584]

杜一(1988－),男,山东聊城人,博士,助理研究员,CCF 专业会员,主要研究领域为信息可视化,人机交互.

郭旦怀(1973－),男,博士,副研究员,CCF 专业会员,主要研究领域为地理信息系统.



陈昕(1982－),女,博士,助理研究员,主要研究领域为计算机图形学.

任磊(1979－),男,博士,副教授,CCF 专业会员,主要研究领域为信息可视化,云制造.

戴国忠(1944－),男,研究员,博士生导师,CCF 高级会员,主要研究领域为人机交互.

© 中国科学院软件研究所 http://www.jos.org.cn