

重庆市软件行业产品发展现状及趋势分析^①

马语晗¹, 郑旭飞¹, 赵欣²

1. 西南大学 计算机与信息科学学院 软件学院, 重庆 400715; 2. 重庆市软件评测中心有限公司, 重庆 400799

摘要: 进入二十一世纪以来, 重庆市软件产业获得了较为快速的发展, 累计完成软件产品测试数量超过 1 万个. 但和全国前 10 省市相比, 软件业务规模仍有一定差距. 为进一步分析当前重庆市整体软件行业产品的基本现状, 本研究通过采集重庆市软件评测中心自 2001 年以来重庆市软件产品测试样本数据, 建立了重庆市软件产品质量数据库. 以数据库为依托, 利用数学统计方法和图表工具从多角度对数据进行统计和分析并搭建数据可视化平台. 结果表明, 软件产品数量、规模和复杂程度总体呈现波动上升趋势; 软件产品可靠性、可移植性、效率、性能等指标测试用例占比提升; 软件产业地区集聚效应凸显、产业生态持续优化.

关键词: 软件产品发展现状及趋势; 数据可视化平台; 软件产品质量数据库

中图分类号: TP31

文献标志码: A

文章编号: 1000-5471(2023)05-0087-08

Analysis on the Development Status and Trend of Software Industry Products in Chongqing

MA Yuhan¹, ZHENG Xufei¹, ZHAO Xin²

1. School of Computer and Information Science & Software, Southwest University, Chongqing 400715, China;

2. Chongqing Software Evaluation & Testing Centre, Chongqing 400799, China

Abstract: Since the beginning of the 21st century, the software industry in Chongqing has achieved relatively rapid development, with more than 10000 software product tests completed. However, compared with the top 10 provinces and cities in China, the scale of software business still lags behind and lacks comprehensive control. In order to further explore the basic status quo of the overall software industry products in Chongqing, this study established a software product quality database in Chongqing by collecting the sample data of software product testing in Chongqing in the past 20 years since 2001. Relying on the database, this study used mathematical statistics methods and chart tools to conduct statistics and analysis of data from a multi-dimensional perspective and built a data visualization platform. The results show that the number, scale, and complexity of software products are on the rise; The proportion of software product reliability, portability, efficiency, performance, and other indicators has increased; The regional agglomeration effect of the software industry is prominent, and the industrial ecology is continuously optimized.

Key words: development status and trend of software products; data visualization platform; software product quality database

① 收稿日期: 2022-10-27

基金项目: 重庆市技术创新与应用发展专项面上项目(cstc2020jcsx-msxmX0148).

作者简介: 马语晗, 硕士研究生, 主要从事软件测试研究.

通信作者: 郑旭飞, 副教授, 硕士研究生导师.

进入二十一世纪,信息技术快速渗透到经济建设和社会生活的各个方面.而软件作为信息社会的物质核心,已成为我国重要的支柱产业之一^[1].软件产业的发展状况是衡量一个国家综合国力的标志之一,也是引领新一轮科技革命的关键力量.其中,重庆市软件产业也保持了快速发展势头,通过两江数字经济产业园、渝北仙桃大数据谷等 4 大重点园区形成了较好的集约效应,培育出了中冶赛迪、长安软件、金算盘、猪八戒网等多家本地知名软件企业.重庆市软件产业整体呈现平稳向好发展态势,收入和利润均保持较快增长.

通过对近 20 年来重庆市软件产品的跟踪统计,发现全市软件产品的数量和质量总体上是在逐步提高的,但当前重庆市软件行业的整体规模仍较小,绝大多数软件企业还是从业人数在 50 人以下的小企业.因此,为探索重庆市软件行业产品的发展现状及趋势,本研究以重庆市软件评测中心收集的近 20 年重庆市软件产品测试数据样本为依托,从第三方软件测评角度,结合软件产品的企业地区分布情况、软件产品类型、软件外部质量特性指标等属性,对重庆市软件行业产品发展现状进行可视化分析;结合软件行业的现有政策背景对统计结果展开深入讨论,探索重庆市软件行业产品未来的发展趋势,为重庆市软件行业发展提供有力的数据支撑和政策参考.

1 数据准备工作

数据准备工作是进行数据分析的基础,主要包括数据采集和数据清洗两个方面^[2].数据的准确性直接影响分析结果的客观性.因此,数据的准备工作十分重要.本研究通过大量的前期调研工作并结合实际研究目的来确定数据采集指标.采集指标主要包括:所测软件产品的基本信息(软件名称、软件版本号、软件类型、测试类型)、所测软件产品的外部质量特性指标(软件总功能数、测试总用例数、功能测试用例数、易用性测试用例数、可靠性测试用例数等)、所测软件产品的企业基本信息(企业名称、企业地址)等.

1.1 数据采集及清洗

采集重庆市软件评测中心自 2001 年以来的 5 005 个重庆市软件产品测试样本,经过筛选后得到数据:软件产品基本信息 3 887 条,软件测试基本信息 3 887 条,软件外部质量特性指标 2 108 条,软件产品特点指标 2 106 条,企业基本信息 1 420 条.其中,软件外部质量特性指标和软件产品特点源自 2009 年—2020 年软件测试技术档案,需编写自动化脚本结合人工操作进行数据抽取,数据采集示意图见图 1.

由于 Python 自动化脚本采集的原始数据类型不一,部分文档存在数据重复采集且格式不规范等现象.因此,本研究对采集的数据进行如下处理:清理异常值、纠正错误数据、清除重复值以及规范数据操作.通过上述步骤将不同数据源中不同格式的数据转换成了统一的格式,并去除错误记录、重复记录,补充遗漏记录,以提高数据质量^[3].除此之外,对于存在异常值的数据需结合测试报告等文档进行充分比对后完成数据核验.

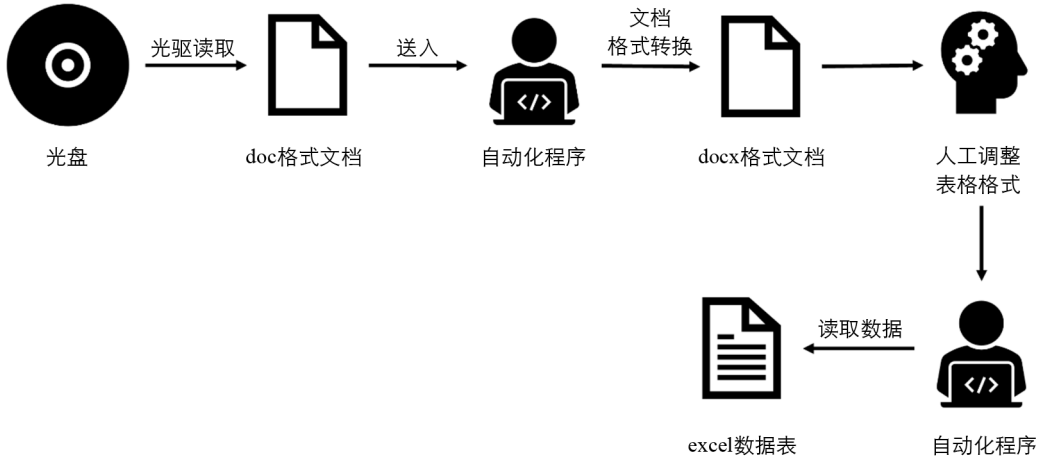


图 1 数据采集示意图

1.2 搭建数据库

为加强对测试样本数据的管理, 并提高管理效率便于开展后续的分析工作, 本研究根据需求构建企业基本信息表、软件产品基本信息表、软件产品测试情况信息表以及软件产品外部质量特性指标表. 在数据库的选取上, 选择 MySQL 数据库. 与其他大型数据库的设置和管理相比, MySQL 数据库复杂程度较低、容易学习、可移植性较强, 便于后续维护^[4].

重庆市软件产品质量数据库的建立是对重庆市软件产品现状的数字化体现, 具有重要的研究意义. 同时, 又极具实践意义, 实现了对测试样本数据的共享, 减少了数据文档的冗余度, 便于对测试样本数据进行集中控制以有效分析重庆市软件行业产品发展现状, 为后续重庆市软件行业产品的相关研究提供参考.

2 软件产品现状及趋势分析

2.1 软件产品发展现状及趋势的多维度分析

2001 年—2021 年采集的软件产品基本信息为 3 887 条, 对数据进行清洗和处理后, 剩余数据 3 879 条. 以年份为横轴, 软件数量为纵轴绘制软件产品类型年度数量分布图(图 2). 由图 2 可知, 2001 年—2021 年, 软件数量整体呈现上升、下降、再次上升趋势, 其中: 2001 年—2005 年和 2010 年—2014 年, 软件数量呈现上升趋势, 涨幅较大; 2006 年—2008 年, 软件数量整体趋于平稳, 保持在 150 个左右; 2015 年, 软件数量显著下降, 但于 2016 年—2021 年有所回升. 总体来看, 软件数量总体呈现波动上升趋势, 软件行业发展方兴未艾.

从分布情况看, 软件产品类型以信息管理系统和应用软件为主, 数量占比约为 71%. 2005 年—2017 年重庆市自主开发的嵌入式软件、特定外设应用软件的市场前景较好^[5], 数量占比共约 10%, 随着对汽摩、电子、装备制造等“硬产业”发展的倾斜度和扶持度增大, 占比将进一步提高. 此外, 财务软件、多媒体应用类软件、办公自动化系统、教学软件持平, 数量占比各约 2%.

结果表明, 重庆市软件行业整体蓬勃发展, 软件产品类型日益增多. 但测试样本数据表明重庆市软件行业整体存在“重硬轻软”的现象, 应促进各方向均衡发展, 重点发展政府管理、民生服务、公共安全等行业的应用软件. 除此之外, 在新兴软件上, 可以重点发展大数据、云计算、人工智能等领域的软件产品^[6].

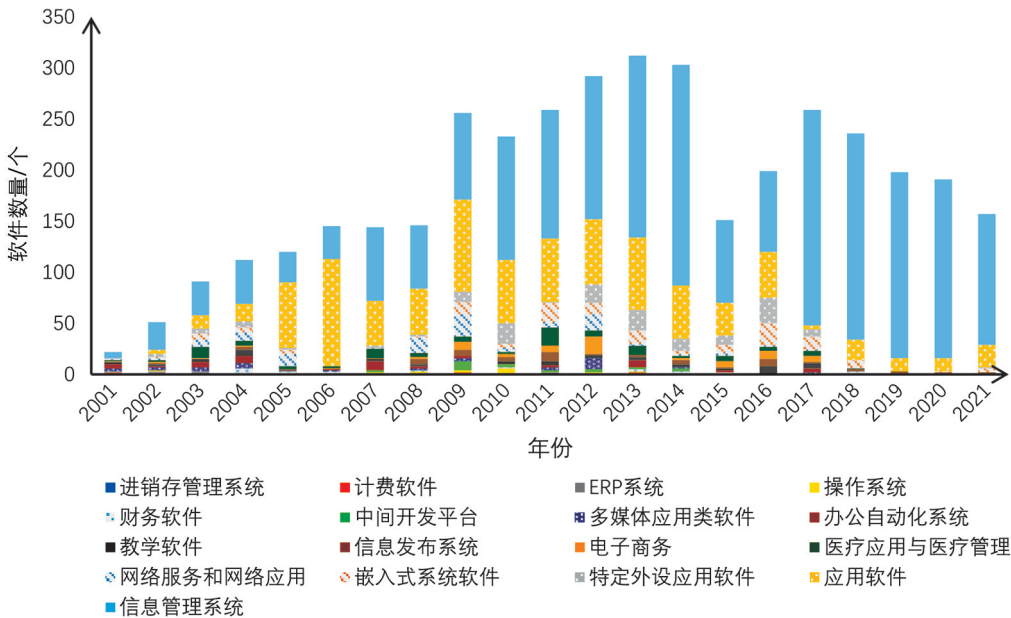


图 2 软件产品类型年度数量分布图

2001 年—2021 年采集的企业地址共 1 420 条, 对数据进行清洗去重且剔除不是重庆的企业地址, 剩余数据 1166 条. 初步统计表明, 测试样本数据中软件产品所属企业大致分布在包含主城区在内的 12 个区县, 且具有较强的集聚效应: 渝北区占 16%, 沙坪坝区和渝中区各占 12%, 北部新区占 14%, 其他地区则主要

分布在丰都县、涪陵区、永川区以及北碚区等区县. 软件产品所属企业地区分布具体情况如图 3 所示.

结果表明, 经过不断发展, 软件产业集约成效突出. 其中, 北部新区作为两江新区的核心区, 软件产业发展迅速; 渝中区依托总部经济和金融中心着力打造外包产业集聚区; 江北区凭借接近城市中心优势吸引部分企业入驻. 此外, 永川、北碚、沙坪坝、南岸以及巴南等地区的企业也利用各自的资源优势促进了软件产业的差异性发展. 总体来看, 产业集群效应初步显现^[7].

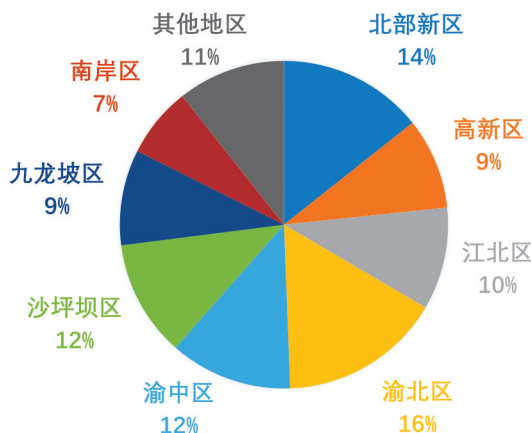


图 3 软件产品所属企业地区分布图

2009 年—2020 年采集的软件外部质量特性指标共 2 108 条, 对数据进行清洗去重以及核验后, 剩余数据 2 049 条. 根据测试时间对软件产品进行升序排序并编号. 以软件产品序列号为横轴, 以软件外部质量特性指标为纵轴绘制散点图(图 4), 其中: FC 表示功能测试用例数; IS 表示信息安全性测试用例数; FT 表示容错性相关测试用例数; RT 表示可靠性测试用例数; TA 表示可移植性测试用例数; EP 表示效率、性能测试用例数; MB 表示可维护性测试用例数; RS 表示运行稳定性测试用例数; CP 表示兼容性测试用例数.

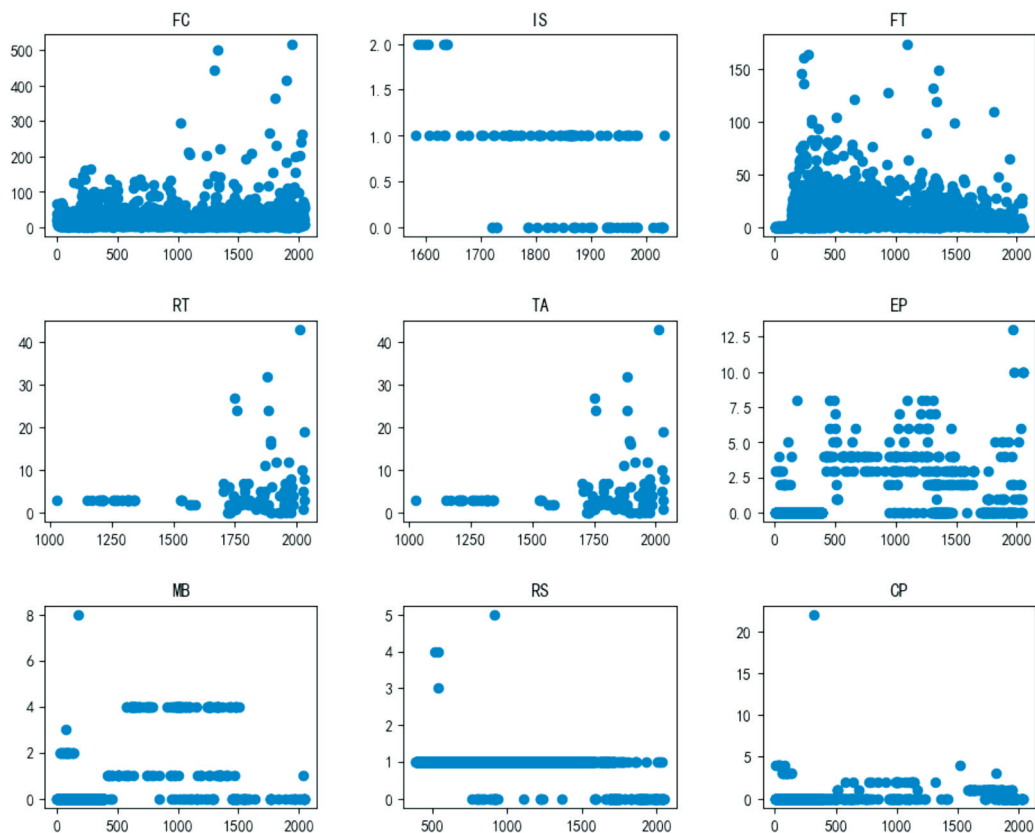


图 4 软件产品外部质量特性指标散点图

由图 4 可知, 2001 年—2020 年, IS, MB, RS, CP 指标数量变化较小. 2014 年—2020 年, 测试样本数据中软件产品的功能点增加, FC 指标呈上升趋势. 2018 年—2020 年, RT, TA, EP 指标数量呈上升趋势. 结果表明, RT 指标与 TA 指标呈正相关, 软件产品性能测试用例数指标的提升与可移植性、可靠性等其他因素有关^[8]. 良好的可移植性可提高软件的生命周期, 是软件产品的能力属性, 与重庆市软件行业环境(软件环境、硬件环境)快速发展密切相关. 软件可靠性测试用例数比重提升, 表明对软件的可靠性达到较高要求、软件中残存错误数较小.

功能点法是可信度高的软件规模度量方法, 基于用户需求来衡量软件的整体规模^[9]. 以年份为横轴, 软件产品平均功能点数量为纵轴绘制软件产品平均功能点数量趋势图(图 5). 由图 5 可知, 2009 年—2020 年, 软件产品平均功能点数量整体保持快速增长态势. 2013 年—2014 年达到低谷, 软件平均功能点数量为 19 个. 2015 年—2018 年, 软件产品平均功能点数量增速较平稳. 2019 年—2020 年呈较大幅度增长. 通过对软件产品平均功能点数量进行统计, 侧面反映了软件产品的规模和复杂程度.

结果表明, 软件规模大小整体呈现逐年上升趋势且保持快速增长态势, 与重庆市政策、区位等因素密不可分. 其中, 重庆市为促软件产业发展, 于 2008 年和 2010 年建成保税区, 吸引大量软件巨头, 致使软件产品规模大幅提升; 2014 年出台相关政策, 经认定的软件企业可按规定享受优惠政策, 导致软件产品数量增多; 自 2019 年以来, 集防疫物资盘点、新冠肺炎疫情实时地图等多功能为一体的软件产品崭露头角, 成为重庆在建工程项目复工复产的“利器”.

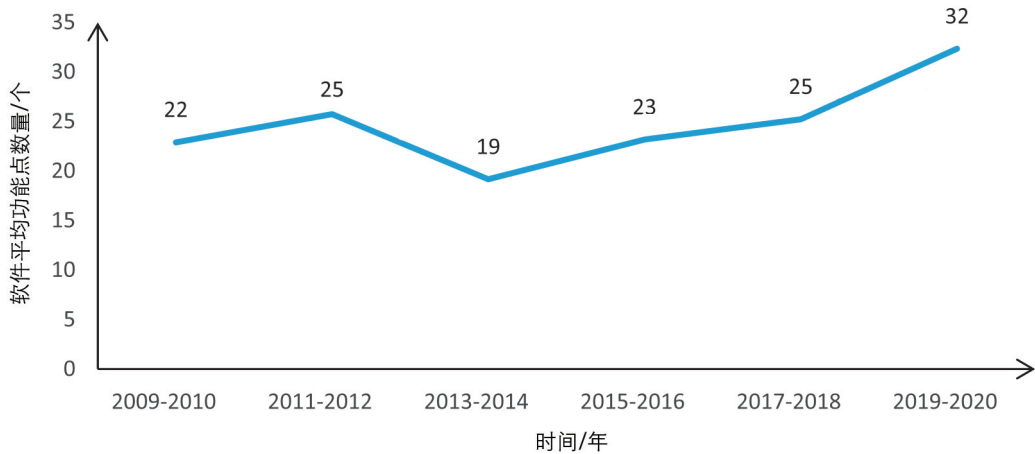


图 5 软件产品平均功能点数量趋势图

2.2 可视化平台的搭建与实现

由于测试样本数量较大且纷杂, 搭建数据可视化平台可以针对同一类型的数据形成不同形式的可视化目标, 具有较强的实用性^[10]. 除此之外, 充分利用可视化技术形成重要信息图表, 将有助于研究人员更好地分析现状及趋势^[11]. 本研究采用 B/S 架构来搭建基于 Boot Strap 框架的 Django 系统. 该系统主要包括主页、数据分析、图表展示 3 个模块, 其中, 主页界面(图 6)展示了软件产品、软件测试类型、软件类型的数量及相关政策和测试标准.

依据不同的展示对象对测试样本数据进行分类后搭建数据分析模块. 该模块主要实现了对数据库中基本信息的查看以及多条件查询功能, 用来辅助研究人员进行数据分析工作. 图 7 和图 8 分别展示了对软件产品基本信息数据的多重条件查询功能以及对软件产品外部质量特性指标数据的管理功能, 便于研究人员在庞大的测试样本数据中筛选出需要的数据并进行管理. 同时, 为满足交互需求, 该模块还增添鼠标悬停、分页展示等操作.



图 6 主页界面



图 7 软件产品基本信息查询界面

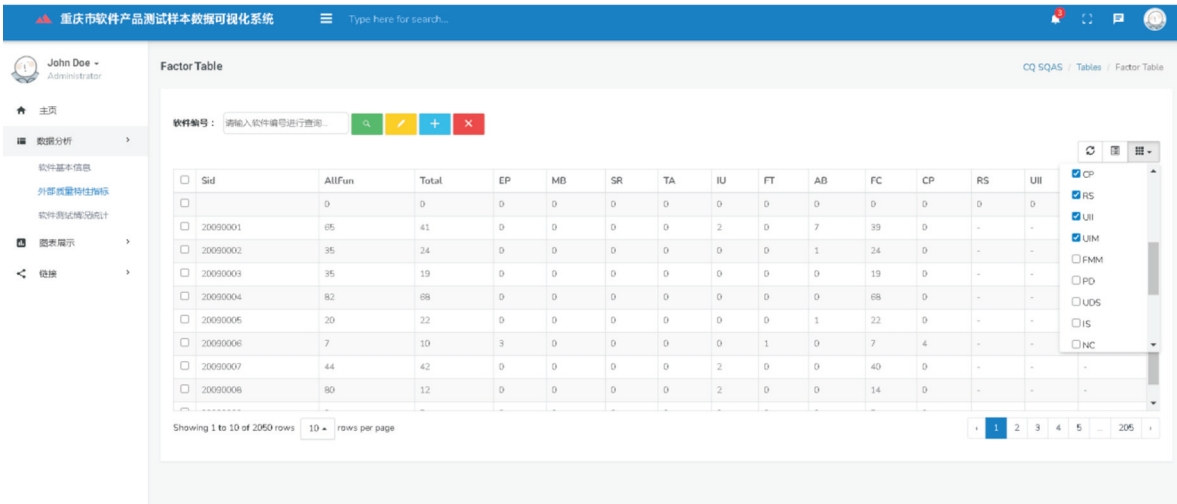


图 8 软件外部质量特性指标管理界面

图 9 所示的 3 个子图均为平台图表展示模块^[12]. 利用文本处理方法对软件产品名称分词, 进行词频统计后绘制词云图(图 9c). 词云图展示软件产品的研究热度, 表明智能化、信息化、大数据、服务、管理等领域的软件产品占比较高. 图 9a 和图 9b 分别展示 2001 年—2021 年软件测试类型分布情况和软件测试类型(按年份)统计结果.

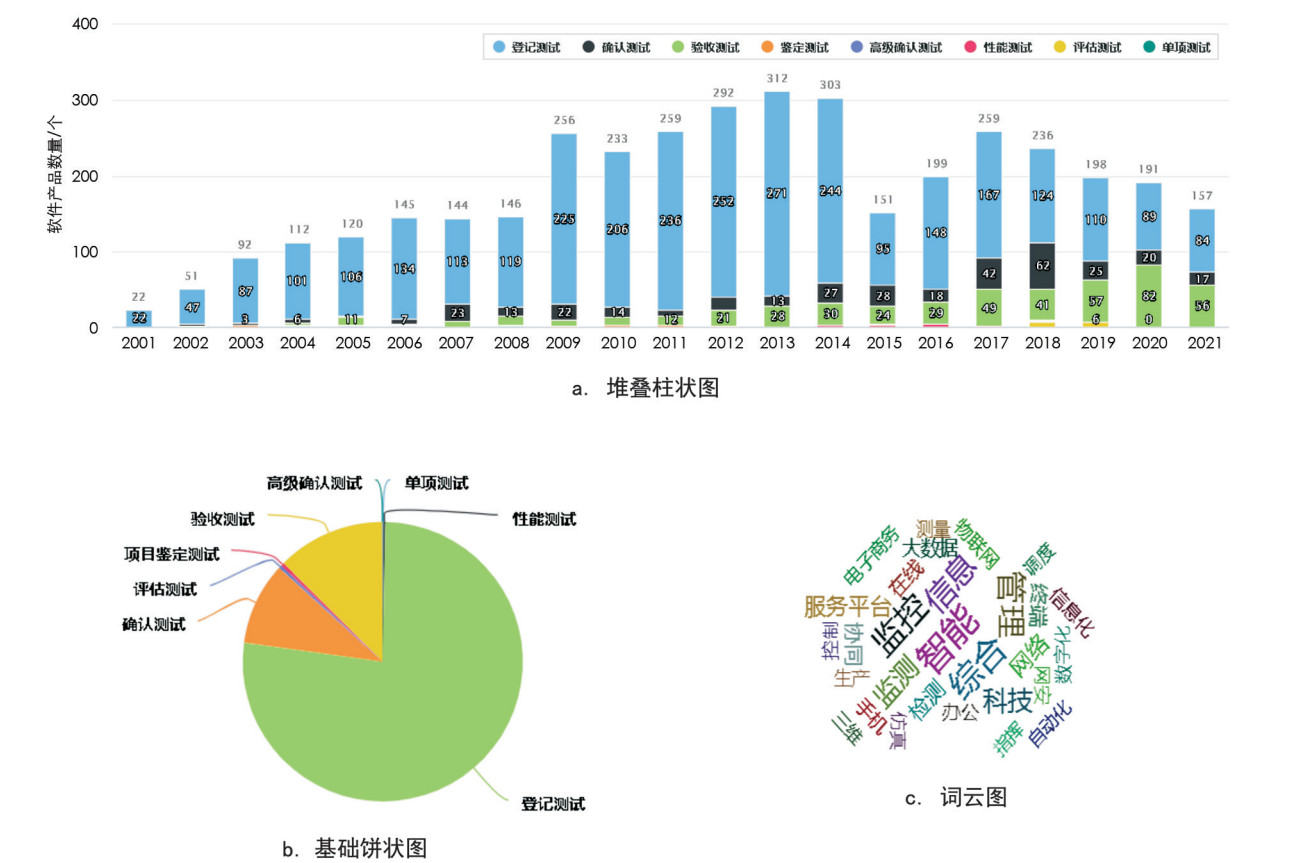


图 9 图表展示模块

基础饼状图表明测试类型主要以登记测试、验收测试和确认测试为主, 每组数据相对于总数的占比分别为 77%, 12.4% 和 9.6%. 堆叠柱状图不仅反映了“总量”与“结构”, 还形象地展示了大分类下各小分类的占比^[13], 表明 2001 年—2014 年随着全市软件行业的快速发展, 所测试的软件产品数量整体呈现平稳上升趋势. 2014 年, 国务院颁布《关于取消和调整一批行政审批项目等事项的决定》后, 软件登记测试数量略有减少, 验收测试和确认测试等大型测试项目比重有较大提升.

3 结论

本研究通过采集重庆市软件评测中心自 2001 年以来近 20 年重庆市软件产品的测试样本数据, 建立了重庆市软件产品质量数据库. 以数据库为依托, 对测试的软件产品所属企业基本信息、软件产品基本信息以及软件外部质量特性指标等数据展开深入分析研究, 利用数学统计方法结合图表工具展现重庆市软件行业产品的发展现状及变化趋势; 搭建数据分析可视化平台, 将统计结果以饼状图、柱状图、词云图、地区分布图等可视化图表形式呈现, 直观展示重庆市软件产品近 20 年发展趋势. 通过数据可视化技术将各类测试数据以及繁杂的统计结果变得简单易懂, 增加决策的准确性^[14].

综上所述, 重庆市软件产业呈现平稳快速发展态势, 取得显著成效; 软件产业地区集聚效应凸显、产业生态持续优化; 软件复杂度和规模度呈平稳上升趋势, 软件产品可靠性、可移植性、效率、性能等指标的

比重提升. 纵向对比, 企业普遍存在“重硬件、轻软件”现象. 现有软件产品在大数据、云计算等新兴领域还处于起步阶段, 创新能力较弱. 还需充分结合重庆市的实际情况, 加速制造业向智能化、服务化、生态化发展的转型, 实现工业软件产业的优化升级, 促进软件产业的快速发展^[15].

参考文献:

- [1] 周鑫贝. 四川省软件产业发展效益研究 [J]. 科技经济市场, 2022(3): 23-25.
- [2] 沈艳红, 张娣. 文献计量分析中的数据准备工作研究 [J]. 图书馆建设, 2012(5): 90-92.
- [3] 陈琳. 我国图书情报领域学术微信公众号建设现状分析研究 [D]. 长春: 东北师范大学, 2021.
- [4] 百度百科团队. MySQL [EB/OL]. (2016-01-25)[2022-10-19]. [https://baike.baidu.com/item/MySQL/471251? fr=Aladdin](https://baike.baidu.com/item/MySQL/471251?fr=Aladdin).
- [5] 寇村图. 重庆集成电路产业与工业软件产业发展现状及对策 [J]. 重庆行政, 2021, 22(1): 102-103.
- [6] 夏元. 重庆离中国软件名城还有多远 [N]. 重庆日报, 2022-01-10(4).
- [7] 罗旺, 孙帅, 吕蕊, 等. “一带一路”下的重庆市软件产业发展研究 [J]. 中外企业家, 2015(22): 18-20.
- [8] 李洁. 软件测试用例设计 [J]. 电脑编程技巧与维护, 2010(4): 17-19, 23.
- [9] 朱明英, 姚文胜, 邢豫, 等. 基于功能点法的软件工作量评估方法研究与应用探讨 [J]. 现代计算机(专业版), 2017(33): 52-55, 64.
- [10] 邢翀. 基于 WEB 的数据可视化系统研究 [J]. 信息记录材料, 2020, 21(9): 132-133.
- [11] 谈叶月. 电力行业数据可视化系统研究与实现 [J]. 电力设备管理, 2022(8): 249-251.
- [12] 张玉叶. 基于 PyEcharts 的数据可视化 [J]. 电脑知识与技术, 2022, 18(2): 24-27.
- [13] 任彦晖. 互联网医疗平台的数据可视化设计研究 [D]. 北京: 北方工业大学, 2022.
- [14] 温丽梅, 梁国豪, 韦统边, 等. 数据可视化研究 [J]. 信息技术与信息化, 2022(5): 164-167.
- [15] 邵珠峰, 赵云, 王晨, 等. 新时期我国工业软件产业发展路径研究 [J]. 中国工程科学, 2022, 24(2): 86-95.

责任编辑 张构