

基于Python语言方法的儿童绘本风格与创作者分析

张诗笛*, 孙志鹏*, 高宁

大连外国语学院软件学院, 辽宁 大连

收稿日期: 2023年4月24日; 录用日期: 2023年6月1日; 发布日期: 2023年6月12日

摘要

在当代社会, 儿童问题是热点问题。儿童图书的品质更是影响着儿童的思维发展, 然而我国绘本的发展并不完善, 同时发现儿童绘本相关的研究却并不多。因此, 本文通过爬取并分析媒体中的部分绘本相关内容, 探究当代绘本存在的问题和不足。本实验使用python技术, 爬取当代读者对于现代绘本的网络评价、微博热点话题的评论以及豆瓣绘本相关的作品名称、作者、内容、类别, 并进行数据分析。同时使用matplotlib数据可视化工具对实验数据进行可视化分析, 并运用正负情感的对比分析, 最终得出如下结论: 一, 我国个别种类绘本的数量存在明显不足, 需要创作者的填充; 二, 绘本的内容质量和画风亟待进一步把控和完善; 三, 不同类型作者写作倾向不同, 写作偏好和写作能力有所差异, 存在较大提升空间。

关键词

爬虫, 绘本, 数据分析, 作者倾向分析, 词云图

Analysis of Children's Picture Book Styles and Creators Based on Python Programming Methods

Shidi Zhang*, Zhipeng Sun*, Ning Gao

School of Software, Dalian University of Foreign Languages, Dalian Liaoning

Received: Apr. 24th, 2023; accepted: Jun. 1st, 2023; published: Jun. 12th, 2023

Abstract

In contemporary society, children's issues are hot topics, and the quality of children's books is

*共一作者。

known to influence the development of children's thinking. However, the development of picture books in China is not yet perfect, and there is a lack of research related to children's picture books. Therefore, this study aimed to explore the problems and shortcomings of contemporary picture books by crawling and analyzing some picture book-related content in the media. Python technology was used in this experiment to crawl and analyze the online reviews of contemporary picture books from social media platforms such as MicroBlog and Douban. The titles, authors, contents, categories, and evaluations of netizens were systematically analyzed. Using the matplotlib data visualization tool, the experimental data were visualized and analyzed. And through a comparison of positive and negative sentiments, the following conclusions were drawn: First, there is a significant lack of certain types of picture books in China, which needs to be addressed by more authors. Second, the content quality and style of picture books require further control and improvement. Third, different types of authors have different writing tendencies and differences in writing preferences and abilities, and there is considerable room for improvement.

Keywords

Web Spyder, Picture Books, Data Analysis, Author Tendency Analysis, Word Cloud Diagram

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当代社会, 读书仍然是人们学习知识的重要途径之一。对于儿童, 绘本则是学习外部知识的重要途径。绘本作为重要的儿童读物, 不仅是讲故事、学知识, 还可以全面帮助孩子建构精神世界, 培养多元智能。绘本是发达国家家庭首选的儿童读物, 国际公认“绘本是适合幼儿阅读的图书”。二十一世纪, 绘本阅读已经成了全世界儿童阅读时尚[1]。

随着国内绘本市场的逐渐扩大, 市面上的绘本数量虽然越来越多, 但是质量参差不齐, 许多绘本内容肤浅、故事平淡、画面粗糙, 难以引起孩子的兴趣和阅读热情[2]。由于绘本研究属于冷门, 其数据的选取也存在一定困难。经过在各大社交媒体上多次调研, 本研究最终找到有利于实验研究的话题来进行讨论。豆瓣平台, 其作为绘本书目筛选的优质选择平台, 可以从中批量获取绘本的名称、作者、简介等数据。

本文主要通过引用微博文章评论、豆瓣中各书目的占比以及不同作者的独特写作倾向, 详细分析了国内绘本在画风、内容、种类等方面的不足, 同时分析了不同性别的作者擅长的绘本类型, 并得出结论。

2. 实验数据与流程

2.1. 实验数据

数据来源于微博平台的评论以及豆瓣平台的绘本详情。根据微博平台中“不雅画面、惊悚画面等画风适合孩子看吗”话题进行绘本画风分析。选取热度第一的文章及其评论作为分析的数据。该文章一共有 1.4 万关注度, 970 条评论以及 195 万的浏览量。根据评论的热度排序, 选取前 800 条评论作为研究的数据。

同时爬取 470 条用于作者倾向分析的数据, 数据来源于豆瓣平台绘本相关版面, 数据包含豆瓣图书分类中儿童读物、亲子读物的排行榜排名靠前的大量书籍相关信息, 信息包括绘本名称、作者、内容、

类别和短评等。

2.2. 实验流程

见图 1, 本次实验数据选择微博平台的《滚远点、不雅画面、惊悚画面……这些绘本真的是给孩子看的吗?》文章及其评论和豆瓣平台的所有亲子阅读绘本的详情。将数据所在的页面地址作为 URL 地址, 通过 request 函数获取网页源代码。上述流程获得正确的用户评论数据和绘本详情信息[3]。得到用户评论数据后, 将数据集中保存在一个 csv 文件中, 以便后续操作。而绘本详情的数据存放在 xlsx 文件中。

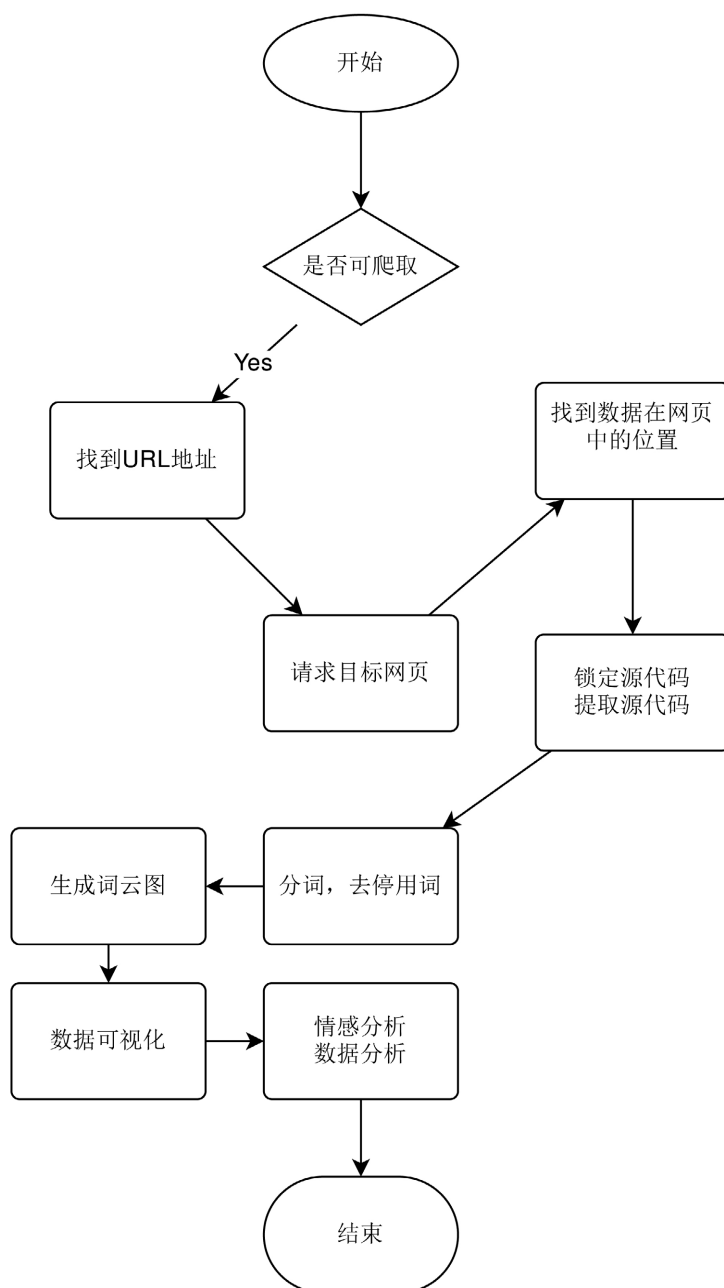


Figure 1. Experimental flowchart

图 1. 实验流程图

首先是对文章和评论的文本数据解析, 进行分词和去停用词操作。在生成词云的时候用到 jieba 分词的第三方库以及 pyecharts 和 WordCloud 生成词云的第三方库。将分割后的词语填写入 txt 文档内, 并将其与中文停用词表中的词语循环进行比较。最终所确定的词语, 是直接分析的数据, 使用 pyecharts 库中的 options 包配置和 WordCloud 包生成词云图。情感分析过程实现方法是通过调用百度 AI 的 API 接口来实现用户评论的情感分析。使用接口前, 在情感定向分析的创建模型中, 上传用于自由训练、生成模型的正向语料和负向语料。然后通过百度 AI 的情感分析方向 API 接口的调用, 提高数据的准确性。在机器筛查后, 进行人工筛查, 进一步精进数据。这一系列操作完成后, 将正向和负向的结果导入两个文档内。

得到绘本详细数据后, 如绘本名称、作者、作者性别、绘本内容、绘本类别等来进行数据分析。使用 pandas 库读取 xlsx 文件, 将通过一系列的函数进行数据的分类收集, 最后使用 matplotlib 库和 pyecharts 库进行数据可视化, 直观地显示到目前为止豆瓣阅读中绘本的详细情况。

3. 实验结果与分析

3.1. 可视化词云图

词云图选取词汇为微博文章《滚远点、不雅画面、惊悚画面...这些绘本真的是给孩子看的吗?》中高频出现的词语。生成三张词云图, 分别关于文章整体内容, 文章负面评论和文章正面评论。

见图 2, 通过爬取文章内容, 生成图 2 文章内容词云图, 可以看出这篇文章主要围绕孩子与绘本内容展开, 其中孩子、绘本、中国、女士、故事、画面等词语频繁出现。这说明该文章是在讨论绘本的画风是否会绝对地影响孩子今后的行为举止。关注这件事较多的可能还是中国女性。然而, 孩子作为新生代的花朵, 他们的认知还在学习阶段。该文章表明绘本质量不佳对孩子会产生一定的影响。绘本作为亲子阅读的代表读物, 在阅读绘本时需要家长对孩子的正向引导。绘本的创作者也更应该以孩子的角度考虑问题, 引导孩子正向发展。绘本画面应尽可能适合每个年龄段的孩子阅读, 尽量避免恐怖、奇怪的画面。



Figure 2. Article content word cloud

图 2. 文章内容词云图

将这篇文章中网友对于“画风是否绝对地影响孩子的认知”的评论生成词云图, 如图 3 和图 4 所示。由图 3 和图 4 可知, 正面评论和负面评论的偏重并不完全相同。正面评论即为各样的画风并不会直接影响孩子今后的认知与行为; 负面评论则为画风的不雅一定会影响孩子的认知。争议点都是这名按摩的男子和对孩子的认知的影响, 正面评论中出现更多的词汇是家长和引导, 大部分网友认为低龄绘本需要家长陪伴阅读, 及时进行正向引导, 以便提高孩子的阅读能力和理解能力, 并且可以避免孩子对绘本的错

合知识和教育内容,也可能代表着家长们希望通过绘本来帮助孩子学习知识和培养好的习惯;科普类、生活类和心理学类的数量相对较少,分别为26本、15本和11本。这可能反映出豆瓣亲子阅读绘本的用户对这些领域的需求不是很高。这些信息对于想要选择适合孩子的绘本或者了解亲子阅读绘本市场的人来说都是有价值的。

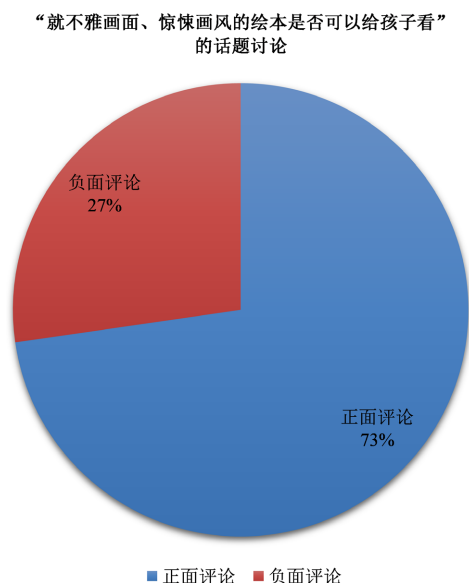


Figure 5. Pie chart of positive and negative reviews
图 5. 正负面评论比例饼状图

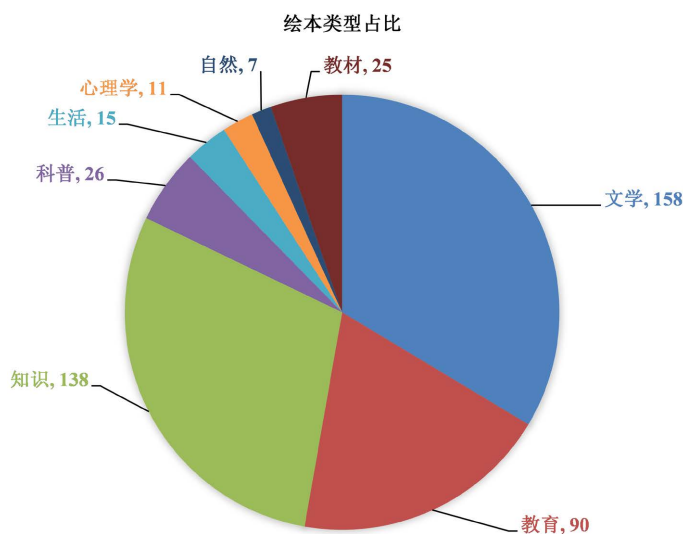


Figure 6. Picture book type ratio pie chart
图 6. 绘本类型占比饼状图

3.3. 绘本作者性别占比分析

根据图 7, 男性作者占比最高, 占比 51.6%。这意味着在豆瓣的亲子阅读绘本中, 男性作者写作的书

籍比女性作者更多。女性作者占比为 31.0%，相对于男性作者而言比例较低，但仍有一定的比例。团体组织占比为 17.4%。这可能指的是由多个作者组成的机构或团队，他们共同出版了一些书籍，因此可以看出在豆瓣亲子阅读绘本中，有相当一部分的书籍由团体组织所创作。这些信息有助于读者了解亲子阅读绘本的创作背景和作者性别分布情况。

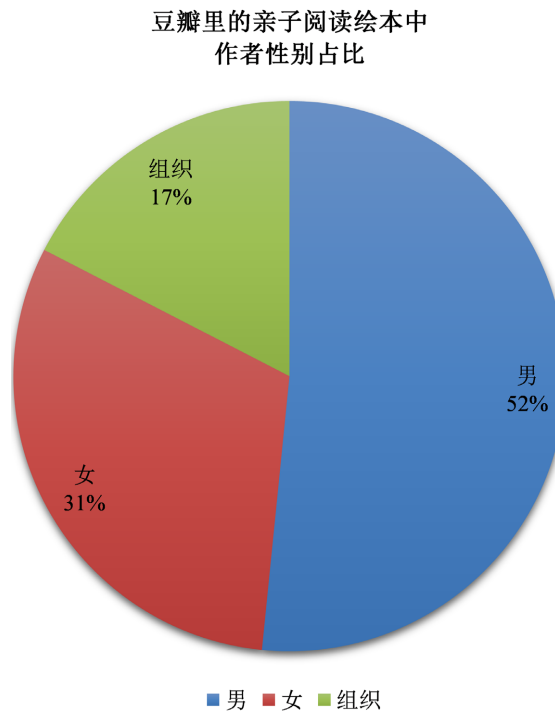


Figure 7. Parent-child reading picture book in Douban reading the author's gender ratio pie chart

图 7. 豆瓣里的亲子阅读绘本中作者性别占比饼状图

3.4. 绘本作者倾向分析

图 8 是根据豆瓣平台相关版面中不同类型作者绘制绘本数量的柱状图。根据图 8，可以看出不同性别的作者对于不同类型的绘本有不同的偏好。文学类绘本是男性作者最擅长的类型，其次是知识类和教育类；而女性作者最擅长的是文学类和教育类绘本，其次是知识类。团队则更多地涉及知识类和教育类绘本。同时，可以发现自然类、心理学类和科普类绘本的数量相对较少。需要注意的是，在本数据中，女性作者的总数量要远远少于男性作者，因此女性作者在每个类别的创作数量都相对较少。这些信息可能有助于绘本创作者了解自己所处团体特长类绘本和缺乏类绘本，也可以帮助出版社和绘本作者更好地了解市场需求。

图 9~11 是根据豆瓣平台相关版面绘制而成的更加详细的作者创作类别情况饼状图。

根据图 9，在所收集的类别中女性作者更偏向于教育类绘本的写作风格，这表明女性作者在教育类绘本方面有独到的见解且更加擅长，同时也更愿意在这方面付出时间。文学类绘本也是女性作者较偏好的类型之一，这可能意味着大部分女性作者也喜欢表达故事和情感，希望用文字和艺术的方式来表达自己的情感。知识类和教材类绘本也受女性作者的关注，占比分别为 24.7% 和 4.8%。这表明女性作者在绘本中也希望传递一些基础知识和技能。

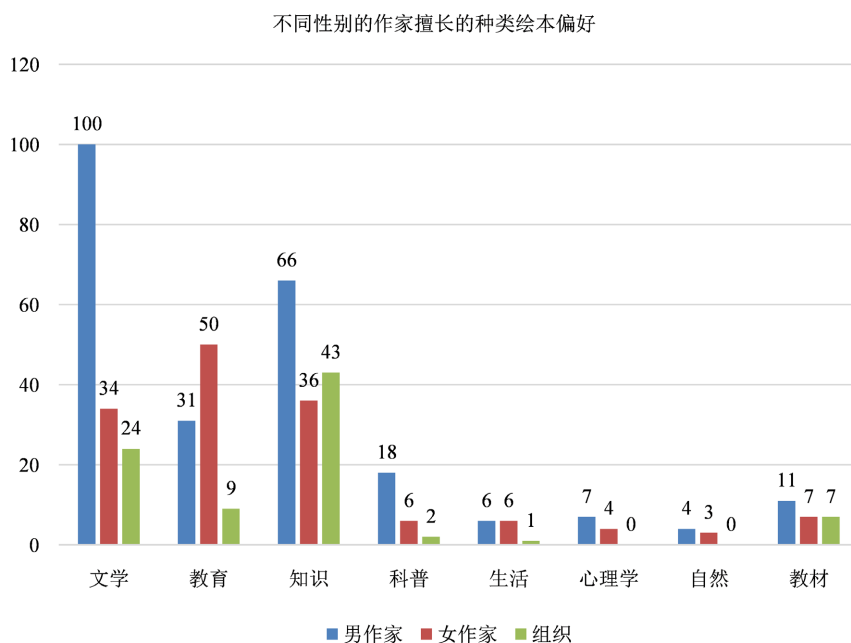


Figure 8. Bar chart of picture book genre preferences by writers of different genders
图 8. 不同性别作家擅长的绘本种类偏好柱状图

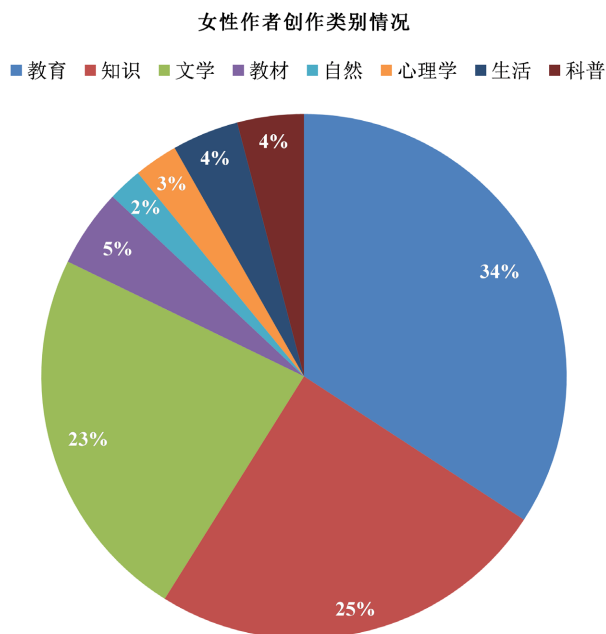


Figure 9. Genres of creation by female authors
图 9. 女性作者创作类别情况

根据图 10, 与女性作者相比, 男性作者在文学类绘本方面占比更高, 而在教育类和知识类绘本方面占比较少。这表明男性作者可能更侧重于在绘本中传递情感和故事, 拥有更多创造和想象的空间, 而女性作者更侧重于在绘本中传递知识和价值观。当然, 这并不是绝对的规律, 每位作者都有自己的创作风格和偏好。

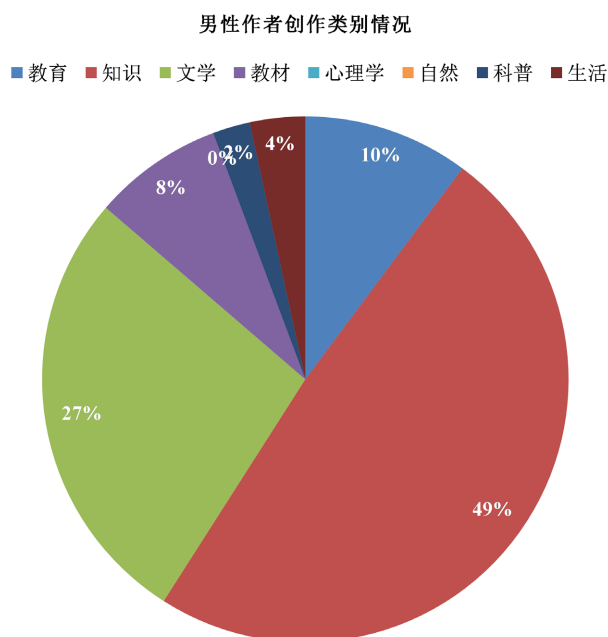


Figure 10. Genres of creation by male authors
图 10. 男性作者创作类别情况

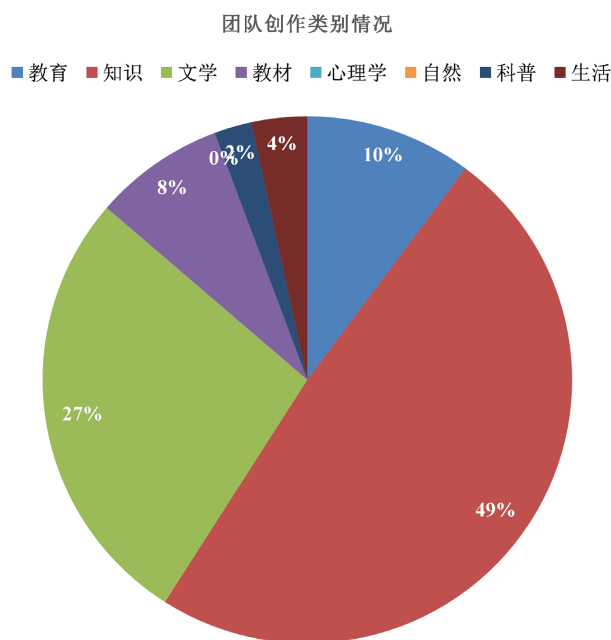


Figure 11. Genres of creation by teams
图 11. 团队创作类别情况

根据图 11, 团队更适合编写知识类的绘本。传授知识和技能需要很多专业人员进行知识的普及和案例分享, 所以更推荐团队去编写知识类绘本。知识的全面讲解需要多人参与和多人讨论。一个知识点的完善, 也是需要做大量的工作。与男性作者相比, 团体在知识类和文学类绘本方面占比更高, 而在教育类和教材类绘本方面占比较少。男性作者则更侧重于在绘本中传递知识。同时, 团体和男性作者在绘本

类型方面的分布比较相似, 这可能反映了团体组织作者在创作时也倾向于采用男性作者的创作风格和方式; 与女性作者相比, 团体组织作者在教育类绘本方面的占比较低。这表明女性作者则更侧重于在绘本中传递知识和教育。同时, 团体和女性作者在创作类型方面有所不同, 这可能反映了团体作者在创作时采用了更加多样化和宽泛的创作风格和方式。

4. 结论

本研究由面向各社交媒体的绘本内容及画风评论出发, 通过数据分析得出:

通过对“绘本画风是否会绝对影响孩子的认知”这一话题的分析, 认为绘本画风不会直接影响孩子的认知的正面的评论为绝大多数, 也因此反映出绘本理解重在家长的正向引导。创作者也无需因舆论的负面导向而限制自己的创作风格, 但其所创作的内容应尽可能适应儿童的身心发展规律。

就“豆瓣绘本”的作者性别以及作品类别数据中分析出女性作者相对于男性作者较少, 应鼓励和培养女性绘本作者, 提高女性作者在绘本创作中的比例; 不同性别作者擅长的绘本种类有所不同, 男性作者大部分专注于知识类、文学类绘本, 女性作者大多关注教育类、文学类绘本, 团队偏好知识类绘本。不同类型的作者可以针对对应的团体选取冷门绘本创作, 以满足不同读者的需求; 绘本类型中, 文学类占比最高, 教育类、知识类占比次之, 而科普、心理学、自然等领域的绘本数量相对较少, 建议创作者在这些领域加强绘本创作和推广, 扩大绘本类型的覆盖范围, 提高绘本的多样性。

5. 分析与讨论

绘本作为儿童认知世界, 探索世界的重要途径, 应当重视其发展与创新, 然而市面上的成品质量良莠不齐, 亟待改进和创新, 经过我们的分析, 为提高国内绘本质量, 需要从以下方面进行改进。

首先, 绘本的内容应当更加丰富和有趣。优质绘本不局限于描述一个简单的故事, 更需要考虑到读者的阅读体验以及对其产生的教育意义。在内容方面, 适当选取更多不同类型的主题, 涉及到自然、历史、人文、科学等方面, 帮助孩子更加全面地了解世界。同时, 在故事情节方面, 需要注意情节的连贯性和紧凑性, 刻画角色形象, 让孩子更清晰地感受到故事中的情感。

其次, 绘本的画面应当更加精美和细腻。绘本的画面是孩子们最直接的感受和认知途径, 优秀的绘本需要拥有绚丽多彩的图画, 用色鲜明、层次分明的画面让孩子们更好地领会到故事的情节和氛围。同时, 优秀的绘本需要注意画面的细节, 例如: 人物形象的表情、场景的描绘等, 以呈现一个更加真实的世界。

最后, 绘本的制作需要注重质量。出版一本优秀的绘本需要整个制作团队的努力。制作团队需要包括作家、插画师、美术设计师、编辑等人员, 他们需要共同协作, 不断推陈出新, 不断地创新和提升, 以确保绘本的精品化。

总之, 提高绘本的质量是一个长期努力的过程, 需要不断地尝试和探索, 同时也需要更多人的参与和支持。只有这样, 我们才能够创作更多更好的绘本作品, 让孩子们在阅读中得到更多的快乐和启迪。

基于以上分析, 国内绘本质量应当在以下几个方面进行提升: 提倡多元化的绘本创作风格和主题, 满足不同读者的需求; 加强在科普、心理学、自然等领域的绘本创作和推广, 扩大绘本类型的覆盖范围, 提高绘本的多样性。提高绘本的质量, 注重绘本故事情节、人物形象、美术表现等方面的创作, 以满足读者对高品质绘本的需求。

基金项目

大连外国语大学 2023 年大学生创新创业训练计划, 项目编号: 202310172A120; “久伴之旅”一新型儿童教育平台研究。

参考文献

- [1] 吕志敏, 张立新. 数字时代儿童绘本创作与出版的思考[J]. 出版广角, 2019(1): 55-57.
- [2] 谭旭东. 中国原创绘本出版发展的观察与思考[J]. 出版广角, 2018(23): 6-9.
- [3] 田雪丽, 郭志斌, 刘梦贤. 基于 Python 的网页数据爬取与可视化分析[J]. 电脑知识与技术, 2022, 18(6): 24-26.