

基于Python网络爬虫技术下的BIM热点话题研究

李晨曦¹, 王冰竹², 李兆辉³

¹北方工业大学理学院, 北京

²浙江财经大学数据科学院, 浙江 杭州

³中国航空油料有限责任公司, 北京分公司第一油库, 北京

收稿日期: 2022年9月26日; 录用日期: 2022年10月16日; 发布日期: 2022年10月28日

摘要

本文通过python的网络爬虫技术, 使用BeautifulSoup库, 对知乎社交平台上对BIM话题的讨论内容进行爬取, 利用jieba库进行分词, 同时进行430份问卷调查, 用SPSS进行分析, 对现阶段欲进入BIM行业和BIM行业的工作人员进行聚类划分, 并用结构方程模型进行BIM技术推广的障碍分析。用可视化图表的方式分析现阶段BIM技术的发展现状, 并对未来其发展进行统计分析建模, 分析其影响, 并对如何更好的推广BIM技术提出建议。

关键词

爬虫, 词频统计, 知乎, BIM

Research on BIM Hot Topics Based on Python Web Crawler Technology

Chenxi Li¹, Bingzhu Wang², Zhaohui Li³

¹School of Science, North China University of Technology, Beijing

²School of Data Science, Zhejiang University of Finance and Economics, Hangzhou Zhejiang

³The First Oil Depot of Beijing Branch, China National Aviation Fuel Group, Beijing

Received: Sep. 26th, 2022; accepted: Oct. 16th, 2022; published: Oct. 28th, 2022

Abstract

Through the web crawler technology of Python, this paper uses Beautiful Soup library to crawl the discussion content of BIM topics on Zhihu social platform, and performs word segmentation by Jieba library. At the same time, 430 questionnaires are conducted and SPSS is used for analysis. At present, the staff who want to enter BIM industry and BIM industry are divided into clusters, and

the barriers of BIM technology promotion are analyzed by structural equation model. This paper analyzes the current development status of BIM technology in the way of visual charts, conducts statistical analysis and modeling of its future development, analyzes its impact, and puts forward suggestions on how to better promote BIM technology.

Keywords

Reptile, Word Frequency Statistics, Zhihu, BIM

Copyright © 2022 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

现阶段网络信息高速发展, 平面二维图纸构建建筑体系, 已无法满足当今高效工作、互联互通的需求。BIM 技术是一种对传统建筑业的颠覆性创新技术, 为连通建筑产业链数据、使得各部门协同工作、提升建筑业信息化水平提供了技术条件[1]。关于 BIM 现阶段发展状况, Li 等[2]分析了 1874 篇关于 BIM 的外文文章, 经研究得出结论: BIM 现阶段研究主要集中在信息系统、3D/nD 模型应用上。崔庆宏等[3]对《工程管理学报》等 10 本期刊上发表过的 BIM 技术论文进行了文本计量分析, 经研究分析, 发现信息化等是现阶段我国 BIM 技术的热点话题。本文基于以上研究, 对第三方讨论社区: 知乎中关于 BIM 技术的相关专区进行讨论内容的爬取, 并用 jieba 库进行分词和停词, 先进性文本分析, 分析现阶段大众关注的热点, 再通过问卷调查, 对使用人员进行聚类分析, 然后建立 SEM 结构方程模型, 分析现阶段 BIM 技术在国内推广面临的障碍。

2. 数据获取

知乎的专题讨论区主要分为三个模块: “讨论” “精华” 和 “等待回答”。“精华” 模块是将数据进行被浏览次数、关注者数量、答案获赞数量降序排序后等汇总的 620 个高质量问答贴和 523 个专栏文章。本文通过爬取精华帖来展开后续分析。工作流程如图 1 所示:

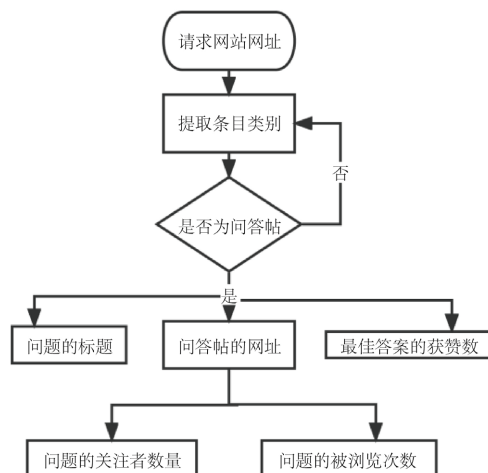


Figure 1. Flowchart of crawler

图 1. 爬虫流程图

使用 BeautifulSoup 库, 先使用 request 指令请求网址, 然后对标题、浏览量、网址、支持度进行爬取, 最后将爬取到的 523 个结果保存到 excel 表格, 用于后续分析。

3. 数据分析

用 SPSS 软件进行建模分析。图 2 对爬取到的数据的 followers、comment、voteup 三项数据进行可视化呈现, 由图可看出, 有两项数据的 followers 极为突出, 分别是① 建筑系学生怎么系统学习 Photoshop、CAD、BIM 之类的软件? ② Revit 有什么奇技淫巧, 让你相见恨晚?

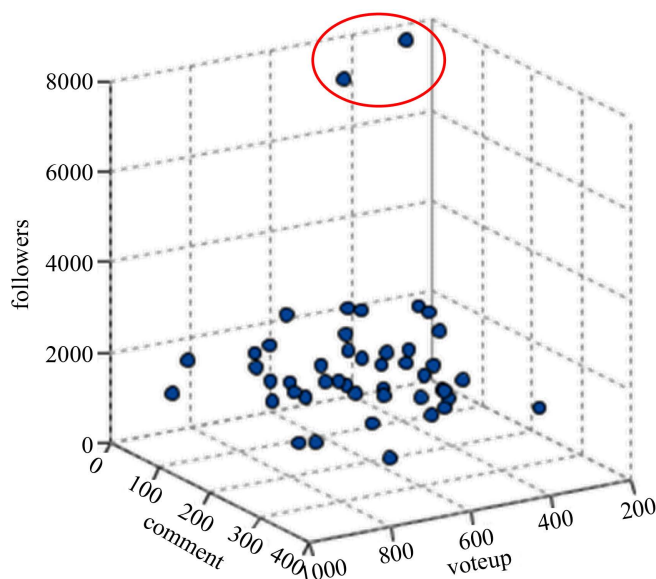


Figure 2. 3D coordinates of data
图 2. 数据三维坐标

利用 python 的 jieba 库进行分词, 并进行统计。排名前 8 的关键词词频排序如表 1 所示

Table 1. Ranking of word segmentation results
表 1. 分词结果排序

排序	关键词	词频	排序	关键词	词频
1	revit	83	5	设计院	32
2	建筑	57	6	工程	23
3	学习	46	7	广联达	19
4	造价	44	8	土木工程	14

由表 1 可知, “revit” 作为关键词出现的词频最多, 为 83 次, 可以看出 Revit 是最热门的 BIM 话题。由此可以看出 Autodesk 公司的 Revit 软件因支持 IFC 标准格式, 并与 Autodesk 公司旗下的 BIM 软件适用广泛, 目前已成为土木工程领域主要的建模软件[4]。通过上述关键词词频统计及排序, 可以分析出现代 BIM 行业的发展趋势。并通过聚类分析, 将上述关键词进行聚类, 分析结果可分为四类: 行业的革新、技术发展、技术应用、技能掌握。聚类结果如表 2 所示:

Table 2. Clustering results
表 2. 聚类结果

聚类	关键词
行业的革新	建筑、专业、造价、设计、行业、设计院、施工、工程
技术发展	Revit、软件、技术、
技术应用	国内、应用、发展、现在、前景
技能掌握	学习、工作、建模、自学

① 行业革新意为将二维图纸模式变成 BIM 3D 建模模式后, 涉及的变化涵盖各行业在 workflows、行业范围、管理模式。BIM 技术通过 3D 模拟技术将平面图纸转换为立体模型, 使得各施工方面能够协同合作, 共同对模型进行更改和查验[5]。随着 BIM 技术的不断发展, 可以推测出, 在未来 BIM 技术必会影响整个建筑、土木行业, 产生革命性的影响。并且, 二维图纸相关行业很有可能被替代取消。从关键词可以看出, BIM 从业人员对未来 BIM 技术的发展存在一定担忧, 存在焦虑。

② 技术发展意味基于 AutoCAD 技术, 实现的广联达造价技术, VR 技术等。主要应用软件有 Revit、archiCAD 等。可以快速访问工程中的几何图形、施工材料、成本、施工进度等信息。施工人员能够快速有效地浏览, 所以能够更加快速有效地制定项目相关决策。并且我们可以从中得知, 包含 revit 的关键词最多, 是当前 BIM 讨论社区里的热点话题。体现出 revit 在 BIM 行业中具有领先的优势地位。

③ 技术发展意为 BIM 发展模式在国内外的比较、BIM 发展现状和未来、施工中应用 BIM 的显著性成果。《建筑业信息化发展纲要》提出[6], BIM 等土木工程行业新科技在工程中的实施, 住建部于 2015 年颁布的《关于推进建筑信息模型应用的指导意见》中同样提出了以国有资金投资为主的大中型建设项目中 BIM 技术的占比达到百分之九十的目标, 以上政策有效加快了 BIM 技术在施工阶段的使用。据上表可知, 所有关键词中, 关于对 BIM 的现状和发展前景的词频出现最高。我们可得知, BIM 技术在工程领域发展得到了广泛的关注。但同时, BIM 技术的发展在我国也具有相当大的挑战。所以, 这也引发行业从业者思考 BIM 在国内外的发展模式的差异性。

④ 技能掌握意为随着 BIM 技术在国内的不断发展进步, 学习 BIM, 使用 BIM 的人员力量正在崛起, 对于 BIM 相关知识的掌握需求变高[7]。人社部于 2019 年 4 月将建筑信息模型技术员列为我国新发展职业。因此, 随着国内制度与技术的不断完善[8], 只会对 BIM 从业者提出更高要求。上表的聚类结果也同样能够反映出 BIM 从业人员的认真刻苦, “自学”在词频统计中出现次数较高。

4. 热点聚焦

利用 python 爬虫结果, 对点赞支持率最高的话题进行降序排序处理, 结果如表 3 所示:

Table 3. Ranking of hot topics
表 3. 热点话题排序

排序	标题
1	BIM 有哪些缺点及推广的障碍?
2	如何系统地学习 BIM?
3	BIM 真的失败了吗?
4	为什么少有设计院能够实现真正的 BIM 设计, 还是死守着 CAD 这样的二维生产工具?

Continued

5	Bim 来了, 传统设计院还能残喘多久?
6	BIM 的引入对工程造价专业意味着什么? 被取代还是共存?
7	为什么 BIM 在国外发展得这么好, 在国内却很难?
8	BIM 会代替现在的 CAD 出图吗?
9	SketchUp 和 Rhino 哪个更适合城市规划建模?
10	BIM 证书工信部和图学会哪个更权威?

表 3 对关注度排名前十的问题进行降序排序, 以研究 BIM 社区化讨论下, 最热点的 BIM 话题。从第一个问题“BIM 有哪些缺点及推广的障碍?”可知, 目前国内关注 BIM 技术的人员最关心的点在于 BIM 技术在于国内的推广障碍, 并且结合第三个问题可知, 国内对 BIM 技术的未来发展存在较大质疑。从第二个问题可知, BIM 新从业者学习能力强, 自我要求高, 关心 BIM 技术的学习问题。后续还有对 BIM 替代传统制图业传统设计院的问题。

5. 人群聚类分析

在此次聚类分析中, 变量为 3 个公因子变量, 分别为使用行为、使用态度、未来期望。基于系统聚类分析的消费人群划分, 是建立在上文的因子分析之后的更进一步的分析, 通过将 430 位被调查者进行二阶聚类, 基于心理因素和行为因素对 BIM 相关人员进行划分, 并根据在 3 个公因子的得分, 进行描述并命名, 为研究不同群体在人文、利益、偏好等方面的特征组合打下坚实的基础。

图 3 所示为通过系统聚类的方法, 将 BIM 相关人员分为 3 类, 其中占比最大的是“潜在使用者”, 占比 49%, 其次是“主体使用者”, 占比 36%, 最后是边缘使用者, 占比 15%。

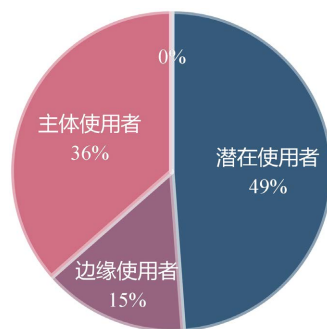


Figure 3. Clustering results

图 3. 聚类结果

对这三类人群进行使用行为、使用态度、未来期望三方面的打分, 结果如表 4 所示:

Table 4. Clustering scores

表 4. 聚类得分情况

聚类	使用行为	使用态度	未来期望
第一类群体(潜在使用者)	1.6	3.16	3.253
第二类群体(主体使用者)	2.4	3.11	3.19
第三类群体(边缘使用者)	1.28	3.7	1.11

第一类群体为潜在使用者, 占比最多, 这类人群对 BIM 技术呈现观望态度, 对其具体使用效果不明所以, 现在较少使用, 但对 BIM 技术未来期望较高。为推广 BIM 技术, 可加大 BIM 技术的宣传, 对其使用效果多加描述, 以吸引更多欲从事 BIM 技术、建筑、土木相关行业人员使用 BIM 技术[9]。

6. 推广障碍分析

本文使用 SEM 结构方程模型, 对推广 BIM 的障碍进行分析。分析结果如图 4 所示:

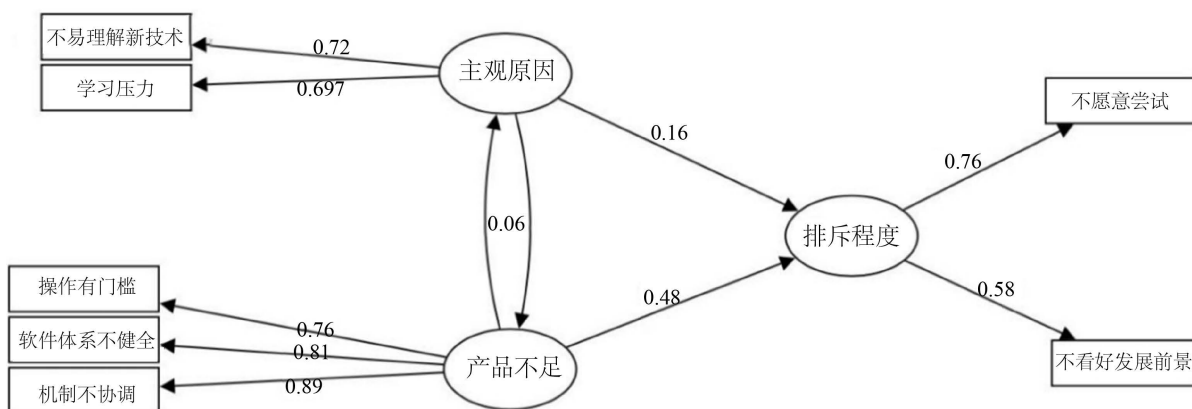


Figure 4. SEM structural equation model

图 4. SEM 结构方程模型

由 SEM 结构方程模型分析结果可知, BIM 应用率低的原因主要是从业人员对 BIM 产生的抵触心理, 因现在大部分企业不具备 BIM 使用资质, 所以自然不会太重视 BIM 人员的培养, 和 BIM 技术的使用。并且, BIM 技术具有技术风险, 影响整个施工工作流程, 所以在当今, 没有切实配套管理机制, BIM 技术难以推广。此外, BIM 技术工种配合不够完善, 缺乏第三方软件支持。在这些方面对 BIM 技术改善, 能够更好的推广 BIM 技术。

7. 结语

本文从第三方社交平台知乎, 通过对 BIM 讨论社区进行网络爬虫, 收集了 BIM 热点话题, 并用 jieba 库进行分词, 统计词频较高的, 并进行聚类分析, 以找到现阶段, BIM 从业者以及欲进入 BIM 行业的新人对这个行业最关注的话题。据分析结果发现, 以 revit 为代表的 BIM 软件是最热门的、讨论度最高的话题。进行聚类分析后, 可将所有分词结果聚为四类, 分别是: 行业的革新、技术发展、技术应用、技能掌握。这 4 类别, 反映了现阶段 BIM 发展的方方面面, 从立体角度分析现阶段 BIM 的发展。并对 BIM 相关群体进行聚类分析, 得出: BIM 技术的潜在使用者最多, 呈观望态度。由 SEM 结构方程模型可知, BIM 应用率低的原因主要是从业人员对 BIM 产生的抵触心理, 因现在大部分企业不具备 BIM 使用资质, 所以自然不会太重视 BIM 人员的培养, 和 BIM 技术的使用。并且, BIM 技术具有技术风险, 影响整个施工工作流程, 所以在当今, 没有切实配套管理机制, 也缺乏相关标准, 应建立跟高层次的应用标准, 以满足 BIM 技术应用的需求。此外, BIM 技术工种配合不够完善, 缺乏第三方软件支持, 在发达国家, BIM 软件已日趋成熟, 但在我国, 还具有相当大的发展空间, 因此, 我们应致力于研发具有中国技术, 自主的软件, 以避免收到国外限制的影响。最后, BIM 技术应用模式单一, 只应用于某一个阶段, 而不是达到互联互通的目的, 我们应不断摸索, 实现应用模式的优化, 争取把 BIM 技术的优点发挥到最大。本文有助于及时把握住现阶段 BIM 发展的动向, 以便今后更好地推广 BIM 技术在国内的发展。

参考文献

- [1] 韩雪, 汪振双, 伊爽. 初步设计阶段的 BIM 造价协同管理模式研究[J]. 工程管理学报, 2019, 33(2): 19-24.
- [2] 郭红领, 潘在怡. BIM 辅助施工管理的模式及流程[J]. 清华大学学报(自然科学版), 2017, 57(10): 1076-1082.
- [3] 崔庆宏, 王广斌, 刘潇, 曹冬平, 路浩. 2008-2017 年国内 BIM 技术研究热点与演进趋势[J]. 科技管理研究, 2019, 39(4): 197-205.
- [4] 马费成, 宋恩梅, 赵一鸣. 信息管理学基础[M]. 武汉: 武汉大学出版社, 2017.
- [5] 李尚昊, 朝乐门. 文本挖掘在中文信息分析中的应用研究述评[J]. 情报科学, 2016, 34(8): 153-159.
- [6] 张永安, 闫瑾. 基于文本挖掘的科技成果转化政策内部结构关系与宏观布局研究[J]. 情报杂志, 2016, 35(2): 44-49.
- [7] 何清华, 钱丽丽, 段运峰, 李永奎. BIM 在国内外应用的现状及障碍研究[J]. 工程管理学报, 2012(1): 12-16.
- [8] 何贵友, 王广斌. 组织变革动能对 BIM 技术采纳的影响机理实证研究[J]. 统计与决策, 2013(7): 97-100.
- [9] 郭博, 赵隽瑞, 孙宇. 社会化问答社区用户行为统计特性及其动力学分析: 以知乎网为例[J]. 数据分析与知识发现, 2018(4): 48-58.