

季珏, 安超, 李波茵, 伍雍涵. 基于文本挖掘的城市园林绿化信息化管理的需求分析方法 [J]. 风景园林, 2019, 26 (8) : 44-47.

## 基于文本挖掘的城市园林绿化信息化管理的需求分析方法

# Method of Demand Forecasting of Urban Landscaping Information Management Based on Text Mining

季珏 安超 李波茵 伍雍涵

Ji Jue, AN Chao, Li Boyin, WU Yonghan

中图分类号: TU985.12

文献标识码: A

文章编号: 1673-1530(2019)08-0044-04

DOI: 10.14085/j.fjyl.2019.08.0044.04

收稿日期: 2019-05-05

修回日期: 2019-07-02

季珏 / 女 / 博士 / 住房和城乡建设部城乡规划管理中心副研究员 / 研究方向为城市园林绿化管理、城市空间结构

Ji Jue, Ph.D., is an associate research fellow in the Urban-Rural Planning Administration Center, Ministry of Housing and Urban-Rural Development. Her research focuses on urban landscaping management and urban spatial structure.

安超 / 男 / 硕士 / 住房和城乡建设部城乡规划管理中心副研究员 / 研究方向为风景名胜管理、城市园林绿化技术管理

AN Chao, Master, is an associate research fellow in the Urban-Rural Planning Administration Center, Ministry of Housing and Urban-Rural Development. His research focuses on urban scenic spots management and urban garden greening technology management.

李波茵 / 女 / 硕士 / 住房和城乡建设部城乡规划管理中心助理研究员 / 研究方向为园林绿化技术管理

Li Boyin, Master, is an assistant research fellow in the Urban-Rural Planning Administration Center, Ministry of Housing and Urban-Rural Development. Her research focuses on landscape greening technology management.

伍雍涵 / 男 / 住房和城乡建设部城乡规划管理中心工程师 / 研究方向为园林绿化技术管理

WU Yonghan is an engineer of the Urban-Rural Planning Administration Center, Ministry of Housing Urban-Rural Development. His research focuses on landscape greening technology management.

**摘要:** 利用信息化的手段辅助园林绿化管理的日常工作能够大幅提高管理效率和管理精准化水平。准确、充分和前沿的用户需求分析,是建立城市园林绿化信息化管理系统和平台顶层架构、功能组成、模块设计等的基础。但由于信息化建设和园林绿化行业管理在专业上存在偏差,使得信息化的需求分析和行业的预测工作沟通壁垒较高。结合大数据的特点,提出一套充分结合用户使用评价、网络新闻、会议记录等文本型大数据,利用文本挖掘方法进行城市园林绿化信息化管理的需求分析和预测的科学方法。分析了三类文本挖掘得出的高频词特点,结合前期对城市园林绿化信息化行业的实际调研,得出了当前中国园林绿化的管理目标在层级转变,办公自动化、物联网感知等前端管理逐步突显的需求特点,为推动园林绿化行业信息化工作提出了建设方向。

**关键词:** 风景园林; 文本挖掘; 信息化管理; 需求分析; 园林绿化

**基金项目:** 国家重点研发计划 (编号 2017YFB0503903)

**Abstract:** Assisting daily landscaping management with the means of informationization can greatly improve the efficiency and precision of management. Accurate, full and frontier user demand analysis is the basis for the establishment of an urban landscaping information management system, and platform top-level architecture, functional composition, and module design. However, due to the specialty deviation in information construction and landscaping industry management, the communication barriers between the informatization demand analysis and industry forecasting are high. According to the characteristics of big data, this paper puts forward a set of scientific methods for demand analysis and forecast of urban landscaping information management with the text mining method, which fully combines the user application evaluation, network news and meeting minutes and other text big data. It analyzes the characteristics of high frequency words obtained from the three types of text mining and, along with the actual investigation of the urban landscaping information industry in the early stage, to reach the characteristics of demand prominent in the front-end management of hierarchical transformation, office automation, Internet of Things awareness in the current management goal of Chinese landscaping, putting forward the direction of construction to promote the information work of landscaping industry

**Keywords:** landscape architecture; text mining; information management; demand forecasting; landscaping

**Fund Item:** The National Key Research and Development Program of China (No. 2017YFB0503903)

在生态文明建设的背景下,城市园林绿化行业迎来了新的发展契机。与此同时,为了满足人民日益增长的物质文化需求,以便民共享、绿色发展等为目标,各地开展了以绿地空间结构优化,绿地景观改造提升等为内容的城市园林绿地的扩容提质工作。城市园林绿化行业的

管理对象日趋复杂,各类绿地资源要素种类和规模不断攀升,绿地系统结构日趋复杂,以信息化手段进行城市园林绿化的行业管理已成为城市园林绿化工作的迫切需求,也是提升园林绿化管理精细化和科学化水平的重要抓手。

目前,不少城市开展了园林绿化的数字化

表 1 城市园林绿化信息化管理的主要内容  
Tab.1 Main contents of urban landscaping information management

工作分类	具体内容	示意
园林绿化数据管理	各类绿地空间信息的管理、校核	公园绿地等其他绿地的空间位置
	各类绿地属性信息的管理、校核	公园管理单位、建设信息等
	园林设施的管理	座椅等园林设施信息管理
	园林绿化资料的归档	日常管理的资料(包含文件、影音、图像等)
园林绿化动态监管	绿线保护监管	侵占绿线的建设行为监管等
	绿规实施监管	绿规的实施效果监测等
	古树名木保护监测	古树名木保护范围划定等
园林绿化辅助决策	行业管理指标的计算和统计	三绿指标、公园服务半径覆盖率等
	行业管理指标对比	横向对比:与国家(生态)园林城市评价指标等的对比; 纵向对比:多期历史对比
	行业管理工作推荐	公园建设地址推荐、病虫害预测预警等

工作,将现有的园林绿化规划设计成果、园林绿化历史资料档案、城市重要园林苗木信息等建立了资料数据库<sup>[1]</sup>。此外,针对这些资源管理,已有学者开展了系统功能的设计和建设工作<sup>[2-3]</sup>,并围绕城市园林绿化的核心重点工作,例如绿地资源调查、古树名木保护,绿线监测等重点工作研发了相应的信息管理手段<sup>[4]</sup>。当前,不少国外学者已经开始关注利用GIS和虚拟现实技术进行城市绿地景观的管理。例如2008年,澳大利亚Payam Ghadirian等利用虚拟现实和GIS相结合的技术对澳大利亚西北部城市Victoria进行景观可视化研究,以期加强决策者和非专家之间的交流,使决策过程更加高效<sup>[5]</sup>。Nyuk & Steve按照新加坡国立大学2005年规划要求构建了校园二维和三维绿化图,将校园分成11个区域,分别计算绿化容积率,进而对校园绿化状况进行评估。在泰国曼谷,政府致力于改善当地的绿化状况<sup>[6]</sup>。Bunvong等收集了曼谷各种绿色基础设施数据,包括普通绿地、街道绿化树、文物保护单位等,借助GIS的分析功能,研究当前树种多样性对将来城市绿地规划中树种的选择带来的影响<sup>[7]</sup>。2001年,Kyushik Oh开发了城市景观信息系统(LSIS)用于管理城市景观信息和分析拟发展项目的视觉影响。借助GIS和图形仿真技术实现强大功能,其底层数据库使用首尔信息系统(UIS)的框架设计,极大

地方方便了两者的数据共享,促进了以计算机为基础的城市景观管理的效益和价值<sup>[8-9]</sup>。可见,目前国内学者在绿地资源管理建库的前端探索较多,而国际则关注管理活动的应用预测等前端需求。此外,在城市规划、地下空间管理等行业,已实现了以信息技术手段辅助行业管理的深层次应用。特别是在城市规划行业,以“统一的空间蓝图”为目标,多地建立了“多规合一”行业管理平台,实现了城市基础空间数据、各类规划数据的统一入库和空间匹配叠加<sup>[10]</sup>,进而为规划审查、项目选址、环境影响评估等深层次横向及纵向行业应用服务。相比较而言,目前城市园林绿化的信息化管理仍处于面向管理者的应用阶段,并未开展纵向的行业推广。

在实地调研中发现,近50%的城市虽然敷设有园林绿化信息化管理系统,但系统处于闲置或夭折状态。其中,行业管理需求、前沿预测与系统设计、建设之间的脱节是造成系统运行效率低下的主要原因。尽管充分的需求沟通是信息化建设的前沿,但受限于沟通者的沟通方式和频率、参与人的管理经验等,导致需求分析的主观性较强,或需求分析结果过于片面,并不能充分发挥和代表行业管理的真实需求或前沿状态。本研究从信息管理系统搭建的最前端需求分析入手,基于文本挖掘的方法探讨一种科学的需求分析方法,

准确反映行业管理的深层次需求和前沿状态,为系统构建和行业管理搭建沟通的桥梁。

## 1 城市园林绿化信息化管理现状

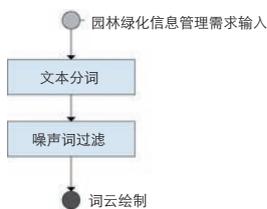
### 1.1 中国园林绿化行业信息化管理的主要内容

结合城市园林绿化管理工作的内容,一般而言,城市园林绿化行业信息化管理的主要内容包括3个方面(表1)。1)园林绿化数据管理。具体是指对城市各类绿地、园林绿化设施等各类城市园林化管理对象的空间、属性和相关资料的信息化管理。2)园林绿化动态监管。具体是指绿线监测、绿规实施监管、古树名木保护监管等。3)园林绿化辅助决策。具体是指针对园林绿化管理工作中公园建设选址、病虫害预警等具体工作开发相关科学决策的辅助信息工具。

### 1.2 城市园林绿化信息化管理平台的发展

20世纪90年代后期,随着信息化、网络化建设进程的加快,各地也相继开发和研制了一些园林管理方面的信息系统。其中一类是基于GIS技术搭建的单一园林绿化资源的管理系统,又以城市植物或树种为主要对象<sup>[11-12]</sup>。这类系统具有树种个体信息及相关管理数据的输入和编辑以及图文展示等功能。另一类系统是集成GIS技术,搭建城市园林管理综合系统<sup>[13]</sup>。这类系统在信息建库的基础上,面向决策应用有了进一步发展。但自2010年后,城市园林绿化信息化管理的相关研究和实践工作出现了一定的停滞。此外,目前专业的园林绿化信息管理平台大多是建立在城市一级,辅助城市园林绿化工作,而对于省或国家层面的平台仍在探索期,这就使得园林绿化的行业管理工作并未在行业体系内开展深度应用。

随着物联网、大数据、云平台的大力发展,特别是城市管理平台“数字化—信息化—智慧化”的发展趋势已越发明朗,未来在新的网络时代、信息时代下,城市园林绿化信息系统的构建如何充分满足管理者、公众的用户需求,同时结合移动终端、物联网终端等多源异构数据,实现园林绿化跨业务、跨部门的精细化管理是未来行业信息系统发展的重点。



1 文本挖掘 workflow 设计  
Design of workflow of text mining

## 2 基于文本挖掘的城市园林绿化管理需求分析

在大数据时代，微博、微信等手机 APP 提供了新型的用户数据，如何充分挖掘这些新型的数据，分析并预测园林绿化行业管理的前沿需求，是城市绿化信息化管理的关键问题。鉴于大部分需求数据，例如用户体验、评价、记录、留言等均为文本型数据，本文作者提出可通过文本挖掘的方法，从文本型大数据中发现高频词，挖掘用户的潜在需求，进而总结形成城市信息化管理的前沿需求和功能组成。

### 2.1 文本挖掘法简介及流程设计

文本数据挖掘 (text mining) 是指从文本数据中抽取有价值的信息和知识的计算机处理技术。它是从大量文本数据中抽取事先未知的、可理解的、最终可用的知识的过程，同时运用这些知识更好地组织信息以便将来参考。目前，文本挖掘方法已在规划方案评比<sup>[14]</sup>、网络舆情监测<sup>[15]</sup>以及目的地印象评价<sup>[16]</sup>等方面展开了应用，利用科学的方法从非结构化的文本数据中，实现了数据挖掘和潜在信息分析。

文本挖掘有 3 个独特的关键步骤：1) 分词，通过分词可以将非结构化数据转化为结构化数据；2) 情感分析，定义情感强度，从文本中挖掘用户的满意度或者观点，即一些词语的出现频率；3) 去噪，即从分出的众多词中剔除干扰的无用的词，提升数据质量。

笔者在 Smart Mining 软件中，设计了文本挖掘的工作流 (图 1)，主要分为 4 个步骤。

1) 园林绿化系统的需求输入。首先，使用文本挖掘目录下的平面文件节点读取相应的文件数据。

2) 对文本进行分词。分词的功能是对中文或者英文文本进行分词，进而从非结构化数据中提取结构信息。分词的结果都是按列输出词语。分词后将在输出表中追加分词、计数和词性 3 个字段。分词字段表示从文本字段中提取出词；计数字段表示按照每个文本，对得到的每个分词出现次数的统计，需注意统计的只是每个词所在文本中出现的次数，并不是所有文本中出现的次数；词性字段显示每个分词的词性，例如名词、动词、结构助词等。

3) 噪声词过滤。噪声词过滤节点的功能是将分词字段中的无关词汇或者不重要的词过滤掉，筛选出重要的词汇。过滤的方法有 5 种，分别是过滤与内置的噪声词库匹配的词汇、与自定义噪声词库匹配的词汇、过滤所有的标点符号、过滤所有的数字和过滤所有字节长度小于 N 的词。这 5 种方法可以只选择一个进行过滤，也可以选择全部进行过滤。过滤后输出表中的列数保持不变，只会过滤掉匹配条件的记录。本研究对于园林绿化需求文本噪声词的过滤采用了以上第 1 种方法。

4) 词云绘制。使用过滤后的关键词及其出现频率绘制词云。

### 2.2 研究数据

本研究对于园林绿化信息化管理需求挖掘分析将采用 3 类数据。

第 1 类为会议记录等文字数据。研究组分别于 2012—2018 年持续 7 年在广东省阳江市、内蒙古鄂尔多斯市、河南省开封市、四川省德阳市、山东省枣庄市等地召开全国数字化园林建设会议，累积了全国多个城市的数字化建设需求及建设经验。

表 2 城市园林绿化信息系统建设试点城市分布

Tab. 2 Distribution of pilot cities applying the urban landscaping information system

所在区域	所在省份及试点城市数量	数量	总计
西部	内蒙古 (3)、四川 (1)、贵州 (1)	5	
中部	黑龙江 (1)、河南 (6)、湖北 (6)、湖南 (1)、安徽 (1)	15	30
东部	山东 (1)、江苏 (1)、浙江 (4)、广东 (4)	10	

第 2 类为用户使用体验评价等文字数据。研究组研发的数字园林信息化系统已在广东省阳江、浙江省丽水、内蒙古鄂尔多斯等全国 30 个城市开展了试点应用 (表 2)，并获取了 30 个城市的验收及使用意见。

据分析，试点城市大多分布于中东部地区，这些城市，特别是沿海的东部城市，经济普遍比较发达，城市整体的数字化基础较好，实际信息化应用水平较高，其意见反馈可充分代表用户对数字化管理的前端需求。

第 3 类数据为来源于各类园林绿化行业主管部门门户网站的新闻发言。本研究搜集了自 2012 年以来鄂尔多斯、阳江等全国 15 个城市园林绿化主管部门对于本城市园林绿化数字化建设系统的经验及未来的需求展望。

## 3 城市园林绿化管理的需求分析

所获得的 3 类数据 (word 文档数据) 作为需求数据，利用大数据挖掘工具 Smart Mining 软件进行文本挖掘分析，试图分析园林绿化主管部门关心的热点问题、系统需求及未来趋势。

笔者整理了文本挖掘软件中的挖掘结果，过滤了若干与需求关键词无关的高频词汇，例如口语化的“我们”“召开”“会议”“热烈欢迎”等。按照软件所选分词的出现频率自行绘制了词云图，图中字体越大表示出现的频率越高。为了便于区分，在自行绘制词云图时，运用了统计学的方法，各分词按照出现频率进行 K-means 聚类，将现有高频词分为 3 类，其中 3 类的分法是参考系统聚类的结果，当现有高频词分为 3 类时，类间差最小 (图 2)。

从词云分布图中可以看出，目前各城市园林主管部门对于信息系统的需求呈现 3 个明



2 园林信息管理需求的词云分布图  
Word cloud distribution diagram of garden information management requirements

显的特征。

### 3.1 园林城市是各城市关注的热点问题

从词云分布图的最高频词分布中可以看出，均是“城市”“园林绿化”“绿地”息息相关的词汇，这反映出利用园林绿化信息化管理是为更好地管理城市园林绿化资源，为城市服务，使城市更生态的初衷。在 2016 年修订的《国家园林城市系列标准》（建城函〔2016〕235 号文，以下简称《标准》）中鼓励城市开展园林绿化管理信息技术应用，在高频词云中，“园林城市”成为最高频类词语，这反映了目前各地以园林城市创建推进各地园林信息系统建设是具有显著成效的。

### 3.2 城市园林绿化信息化管理的目标在层进转变

从词云分布图中可以看出，在第二个圈层结构中，与系统内容相关的高频词为“数据”“数据库”“信息管理”“信息技术”“信息提取”。这间接反映出目前利用信息系统实现对城市园林绿化信息的信息化建库管理是所有城市的基本需求。

而“监督管理、管理手段、科学决策”则表达除信息建库管理之外，各地利用信息系统辅助工作的主要工作内容。此外，在文本挖掘中，出现了 5 个高频词：“数字化”“信息化”“规范化”“精细化”“智能化”。

数字化是所有管理对象的一个电子化入

库的过程；信息化是对数字化的进一步发展，不仅完成电子入库，还包含信息之间的交换和共享等关键技术；规范化是在信息化的基础上，形成自身顺畅的管理业务流程；精细化是从管理手段到管理对象的精细化革新；智能化则是在此基础上，利用各种多源异构数据实现更高效、便捷、实时的管理。从以上 5 个概念的内涵来分析，反映了目前各地园林绿化管理目标在层进的转变，园林绿化的管理需求由单纯的数字化、信息化向规范化、精细化、智能化的前沿进一步转变。

### 3.3 城市园林绿化信息化管理需求由数字化、标准化向智能化转变

分布于最外圈层的词汇，虽然频率相对较低，但却能反映出个别城市在应用领域的若干前沿想法，信息化管理的需求在数字化和标准化的前提下，进一步结合物联网等新技术，向智慧化转变。具体如下。

1) 继续完善信息平台的功能属性和业务推广应用工作。

从高频词来分析，“数据共享”“查询统计”“更新”“预警”等高频词的出现，反映了各管理部门希望信息系统在功能上进行完善，不仅满足数据收集、查询统计等基本功能，更希望对于信息系统的标准能有规范的顶层设计，同时对于系统的动态性和实时性，预测预警的功能有了更高的需求。

此外，还可以看出，“技术人员”“专业”“推广”“培训”等高频词的出现，反映了各地对于利用信息技术带动整体专业管理水平和培养专业业务团队寄予了厚望。

2) 增加物联网技术在园林绿化信息化领域的应用。从词云分析来看，多个地区提到了“探测器”“二维码”“传感器”“移动终端”等大数据获取的技术渠道。这充分反映了各地在信息系统构建时，希望与数字城市、智慧城市相结合的需求，尤其是引入物联网技术，使城市绿地的信息化管理能够并入到统一的城市信息化管理中，实现管理的智能化。结合实际调研，可研发移动终端（二维码、探头、探测器等）与信息化系统的结合，针对病虫害预警、智能灌溉、智能施肥、二维码管理等方面，进一步丰富系统辅助决策模块的功能。

### 参考文献 (References):

[1] 旭晖. 园林植物图文数据库管理系统的研究 [J]. 中国园林, 1997 (5) : 57-58.

[2] 宋丽萍, 朱伟华, 丁少江, 等. 深圳城市绿化管理信息系统的设计 [J]. 南京林业大学学报 (自然科学版), 2003, 27 (1) : 59-62.

[3] 李倩, 杨宏宇, 尚二萍. 城市园林绿化数字化建设标准研究初探 [J]. 风景园林, 2014, 21 (4) : 37-41.

[4] 张晓军, 师卫华, 许士翔, 等. 城市园林绿化信息化管理与辅助决策关键技术研究与应用 [J]. 建设科技, 2016 (7) : 104-105.

[5] PAYAM G, IAN D B. Integration of Augmented Reality and GIS: A New Approach to Realistic Landscape Visualization[J]. Landscape and Urban Planning, 2008, 86(3-4): 226-232.

[6] NYUK H W, STEVE K J. GIS-based Greenery Evaluation on Campus Master Plan[J]. Landscape and Urban Planning, 2008, 84(2): 166-182.

[7] BUNVONG T, LADAWAN P, ROGER K, et al. Urban Green Space, Street Tree and Heritage Large Tree Assessment in Bangkok, Thailand[J]. Urban Forestry & Urban Greening, 2008(7): 219-229.

[8] KYUSHIK O, SEUNGHYUN J. Assessing the Spatial Gistribution of Urban Parks Using GIS[J]. Landscape and Urban Planning, 2007,82(1-2): 25-32.

[9] KYUSHIK O. Landscape Information System: A GIS Approach to Managingurban Development[J]. Landscape and Urban Planning, 2001, 54(1): 81-91.

[10] 张鹏程, 杨梅, 王明省, 等. 基于时空信息云平台的“多规合一”应用系统设计 [J]. 地理空间信息, 2019 (3) : 4-7+9.

[11] 蓝增全. 城市绿化树种信息系统的研究: 对昆明市城市绿化树种选择提供信息支持的实例 [J]. 西南大学学报, 2002, 22 (1) : 29-38.

[12] 王谨, 唐乐尘, 王良睦, 等. 厦门市园林植物管理信息系统 [J]. 中国园林, 2001, 17 (1) : 65-67.

[13] 王霓虹, 范艳芳, 周洪泽. 基于 GIS、RS 的城市局域绿地评价及系统实现 [J]. 东北林业大学学报, 2005, 33 (5) : 75-76.

[14] 伍毅敏. 基于文本挖掘的两版北京城市总体规划公众意见对比分析 [J]. 北京规划建设, 2018 (1) : 87-94.

[15] 黄晓斌, 赵超. 文本挖掘在网络舆情信息分析中的应用 [J]. 情报科学, 2009 (1) : 94-99.

[16] 王媛, 许鑫, 冯学钢, 等. 基于文本挖掘的古镇旅游形象感知研究: 以朱家角为例 [J]. 旅游科学, 2013 (5) : 86-95.

### 图表来源 (Sources of Figures and Tables):

图 1~2 由作者自绘; 表 1 由作者整理自绘; 表 2 试点城市数量来自前期研究团队系统建设实际经验。

(编辑 / 王一兰)