

# 文本分析在规划公众参与中的应用研究

——以“济南城市发展战略规划”之愿景填写分析为例

赵虎<sup>1</sup> 李飞<sup>2</sup> 陈宇<sup>3</sup> (1. 山东建筑大学 建筑城规学院, 山东 济南, 250101; 2. 山东建大建筑规划设计研究院, 山东 济南, 250000; 3. 中国城市规划设计研究院 村镇所, 北京, 100835)

**【摘要】**伴随着网络信息技术的升级和个人手持信息终端设备的普及化, 市民参与城市规划公众参与的途径越来越丰富, 于是越来越多的文本信息成为今天规划公众参与的产物, 而如何有效处理这些众多的文本信息成为规划师亟需面对的问题。本研究首先界定了公众参与中文本分析的类型和需求, 其次提出了海量文本分析的五步技术框架, 最后以“济南城市发展战略规划”的城市愿景填写分析为案例并给出相应的分析解读。通过运用 ROST-CM 软件进行词频和语义网络分析发现, 在济南城市发展战略规划中, 市民对城市愿景的首要诉求是“宜居”, 相关词群主要集中在生态环境、历史文化、交通服务、产业创新、区域定位和地域特色等领域, 并结合关联度分析给出不同的愿景诉求等级。

**【关键词】**文本分析; 城市规划; 公众参与; 城市愿景; 济南

**【中图分类号】**TU984 **【文献标识码】**A

## 0 引言

随着大数据的时代到来, 城市规划编制分析的技术和手段也开始升级。规划师通过相关软件对手机信令数据、社交媒体数据、车载 GPS 数据、公交卡数据等大数据类型进行分析, 获得使用者的时空活动轨迹及特征<sup>[1]</sup>, 能有效地提升城市研究和规划成果的科学性和针对性。同时, 还有一种数据形式——海量文本数据<sup>①</sup>, 也开始受到越来越多的关注。文本是语言的表达, 是基于字符串数据而有别于数值的另一种数据类型<sup>[2]</sup>。近年来, 有些学者利用微博等社交媒体上发布的信息进行文本分析, 获取体验者对城市空间或者环境的满意程度<sup>[3]</sup>; 有些学者分析政府政策文件、领导人讲话以及新闻报道中的文本特征, 挖掘其中隐含的城市发展与规划的有效信息<sup>[4-5]</sup>。另外, 也有人通过某一地域相关诗词等文学作品文本的分析, 挖掘基地蕴含的场所精神和人文特征, 从而对规划设计方案提供参考依据<sup>[6]</sup>。

可以发现, 目前有关文本分析应用于规划公众参与的研究并不多, 不过在未来, 文本分析, 特别是

海量文本分析技术在其中会发挥越来越重要的作用。公众参与是城乡规划及至国土空间规划不可或缺的一部分, 无论规划体系如何变化, 社会对公众参与的需求只能越来越高。而伴随着网络信息技术的升级和个人手持信息终端设备的普及化, 越来越多的人会以非正规的形式参与到规划的公众参与活动中来, 他们可以在政府设定的网页或者开发的专用 APP 上采用直接留言或者规定答题的形式进行意愿表达。因此, 越来越多的文本信息成为今天规划公众参与的产物, 而如何有效处理这些众多的文本信息成为规划师亟需面对的问题。

本文以“济南城市发展战略规划”中的城市愿景文本信息分析为例进行研究, 首先界定了公众参与中文本分析的类型和需求, 其次给出了文本分析的技术框架, 最后对济南的案例进行应用并给出相应的分析解读。本研究不仅有助于推进完善新的分析方法在城市规划公众参与中的应用, 更有助于城市规划方案科学性和针对性的提升, 还有助于深度公民社会的培育和形成。

## 1 城市规划公众参与中的文本类型

文本是语言的表达, 是基于字符串数据而有别于数值的另一种数据类型。根据文本自身呈现出

基金项目: 国家自然科学基金项目(51878393)

的文体格式、篇幅字数等条件,可将其分为三类,分别是词汇型文本、短句型文本和篇章型文本<sup>[7]</sup>。三种文本的定义和具体特征见表 1,其数据可以来源于城市规划编制的各个部分。

表 1 文本数据的类型和特点

文本类型	定义	特点	来源
词汇型	由一个或者多个词汇组成的短文本	内容很少,文本特征明显,字数少,数量一般在几个或者十几个左右。	城市性质和职能(交通中心、商业中心、科技中心等)、城市发展目标。
短句型	由 1-2 句话组成的短文本	内容较少,文本特征明显,字数较少,数量多在 10 到 100 字之间。	论坛留言及回复、咨询建议、意见及反馈、手机短信、聊天记录等。
篇章型	由几个章节或者多个段落组成的长文本	内容较为丰富,文本特征较为明显,语序和表达规范,字数多,在几百至几千甚至上万字不等。	政策文件、政府工作报告、领导人讲话、会议纪要及规划文本等。

资料来源:在参考文献[7]的基础上整理而来

根据《城乡规划法》的规定,城市规划工作可以分为规划编制、规划实施、规划修改和规划监督四个部分,各个部分均有公众参与的活动出现,只是活动的形式、对象和内容上有所差别。目前规划公众参与的方式主要有座谈会、听证会、专家论证会、专家评审会、民意调查、市长热线、公众接待日、公示和城市规划展览系统等<sup>[8]</sup>,产生的文件基本上可以分为两类,一类是会议纪要型,一类是意见汇总型。对于会议形式的活动,如座谈会、听证会、专家论证会、专家评审会,最后产生的文件应该是会议纪要,属于篇章型长文本。而对于广泛征求意见的民意调查、公众接待日、市长热线等活动产生的文件应该是意见汇总或者统计表,其文本类型多是词

汇型和短句型的短文本,但对于现场不回收,并采取网络邮件反馈方式的活动而言,其产生的文件也可能是篇章型的长文本(表 2)。

本文着重研究的是民意调查产生的文件,其文本类型以词汇型和短句型为主,属于短文本。为什么要研究这一类型的文本?综合起来有两个原因:第一,短文本分析是长文本分析的基础,二者分析的基本流程是共通的,都需要化解成最基本的词汇后,再进行相关统计和关联分析。第二,短文本分析时,能将语境的关联影响程度降到最低。因为中国文字释义的丰富性,在长文本中,同一个词如果语境不同,其表达的意义可能南辕北辙,而短文本的分析将会减少类似问题对分析结果的干扰。

表 2 公众参与环节与文本类型对应一览表

序号	公众参与活动类型	参与内容	成果文件	主要文本类型
1	座谈会	随机抽取一部分有代表性市民参加讨论,主要听取他们的意见和建议。	会议纪要	篇章型
2	听证会	抽取部分市民(规划地段内利害关系人)、相关专家或单位听取其对方案的意见或者建议。	会议纪要	篇章型
3	专家论证/评审会	组织相关专家对规划方案进行技术论证和评审。	会议纪要	篇章型
4	民意调查	通过大范围的发放问卷或者网络问卷的形式收集市民的意见和建议。	意见汇总	词汇型 短句型
5	市长热线	全体市民可以通过电话、APP、微博、信箱等相关方式表达对规划方案的意见。	意见汇总	词汇型 短句型
6	公众接待日	组织相关工作人员进社区或街道的形式听取市民对城市建设和规划方案的意见。	意见汇总	词汇型 短句型
7	城市规划展览系统	市民通过实地参观城市规划展览馆中展示的规划方案,并对规划方案进行现场评价。	留言本或 意见反馈表	词汇型 短句型
8	公示	将规划方案在官网或者指定场所对市民进行公示,市民可通过各种形式对规划方案进行评价。	网络留言或 信件	词汇型 短句型 篇章型

资料来源:在参考文献[7]的基础上整理而来

## 2 文本分析的技术流程解析

文本是记录人们的思想、观点和情感的载体,对文本进行系统分析挖掘,能够了解文本制定者所隐含的价值观、意图和态度等。城市规划公众参与过程中产生的三种类型文本虽有差异,但是文本分析的核心技术框架可以是一致的。通常情况下分为5个步骤,文本采集、文本预处理、文本分析、可视化与解读、多元验证<sup>[7]</sup>,见图1。

### 2.1 文本采集

文本采集是实现文本分析工作的前提,需要结合公众参与的不同方式而设定相应的采集方式,通常有三种方式。第一种是手工录入采集,这种采集方式针对的是纸质调查问卷或表格,还有电话采访。第二种是从管理后台直接提取,这种方式针对的是网络问卷或者调查表的填写。第三种是通过软件进行网络数据爬取,针对的是政府信箱收集的信件或者网站公示规划成果后的大众评论。这里需要说明的是,公众参与中文本分析的数据类型宜称之为海量数据,而不宜为大数据。这是因为大数据的采集方式通常是批量式的网络爬取,但文本数据里面还有手工录入采集。

### 2.2 文本预处理

文本预处理是文本分析的必要基础,需要将采集到的文本按照软件分析的格式要求进行规范处理,包括格式转换、分词和过滤三步。第一步格式转换是要将 xls、doc 等形式的文件转换成适应软件分析的 txt 格式;第二步分词尤为重要,软件依据自带的词库会对 txt 格式的文件进行分词处理,从而将篇章型、短句型和词汇型的文本分成若干个独立的词汇;第三步是过滤,主要是针对短句和篇章型文本数据,其主要目的是对分词后的词组进行筛查,按照分析结果的需要对软件中的过滤词表进行补充,将与分析无关的词组或对分析结果产生干扰的词组过滤掉。

### 2.3 文本分析

结合公众参与的文本特性和成果需求,常用的文本分析内容包括词频分析和语义网络分析两个方面。词频分析是一种基本的文本分析方法,其原理是统计文本中每个词出现的次数,通过“统计词频”来识别海量文本中隐含的高频词汇,进而可以推测文本所表达的主要信息。而语义网络主要用

于描述词汇的状态及其间的关系,生成的语义网络图是由结点和结点之间的连线组成,结点表示高频词汇,连线表示它们之间的关系。在本研究中通过“中心度”<sup>②</sup>、“关联度”<sup>③</sup>来进行具体的量化分析,最终呈现出的各个关键词之间的关系网络。词频分析和语义网络分析均可使用 ROST-CM 软件<sup>④</sup>实现操作。

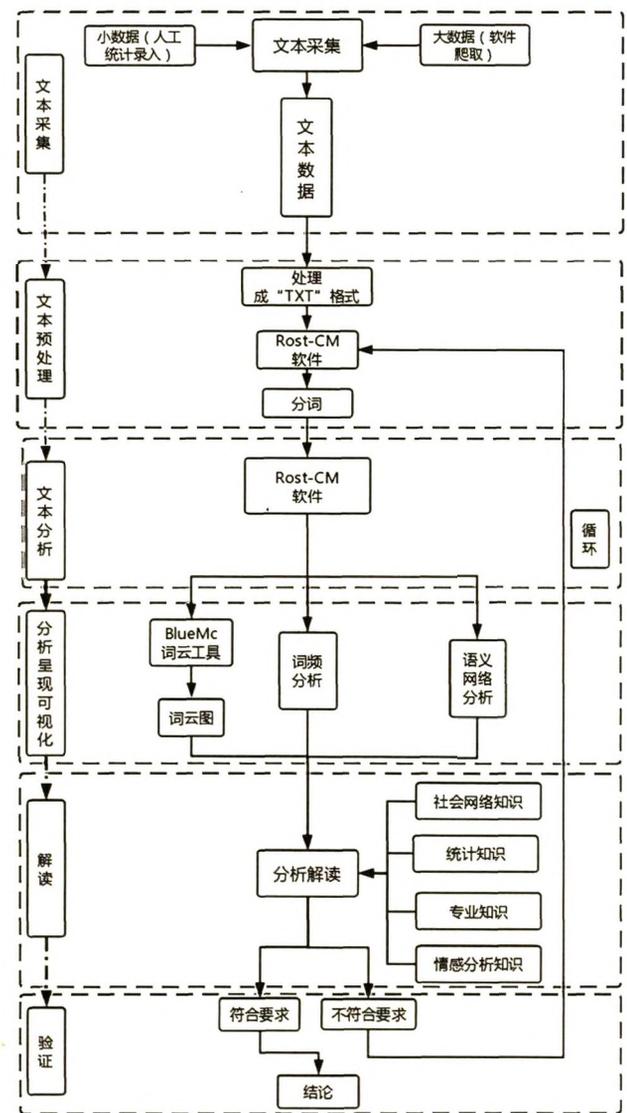


图1 文本分析的技术流程图

资料来源:在参考文献[7]的基础上整理而来

### 2.4 结果可视化及解读

可视化是为了更加直观地呈现分析结构,使文本分析结果以图形的方式呈现出来。其中词频分析可视化的结果多是词云图,可借助 BLUE-MC 词云工具<sup>⑤</sup>实现这一目的。语义网络分析结果能够通

过 ROST-CM 软件生成直观的语义网络图,不仅能显示高频词之间的分布关系,还能显示出两个词汇之间的连结次数(关联度)。同时,分析者需要对统计表和分析图进行专业解读,在这里不仅需要统计学、几何学和社会网络学<sup>[9]</sup>等方面的知识,更需要依赖城市规划专业的工作经验,才能对分析结果作出科学合理的判断解读。

### 2.5 多元验证

考虑到文本分析存在精度上不足的缺陷,所以在技术流程中需要加入人工复检的环节对分析结果进行抽样验证。以词频统计为例,随机抽出 1-2 个词汇进行人工复检,统计其出现的次数和频率,与软件分析的结果进行较核验证。如果偏差太多,则需要返回到前面的某个环节重新调整和改进,然后再进行新一轮的文本分析及验证。在这里称为多元验证,是希望能有更多类型的支持数据可以验证相关结论,以提升文本分析的正确率。

## 3 济南城市发展战略规划应用分析

愿景分析是“济南城市发展战略规划”研究的重要内容之一,是城市居民共同达成的发展共识,在分析制定的过程中需要采集广大市民的意见和建议。由此,“济南城市发展战略规划”研究团队在市民问卷中设计了相关问题,并于 2018 年 6 月 28 日至 7 月 28 日间通过政府机构和专业问卷网站展开广泛调查填写活动(图 2)。

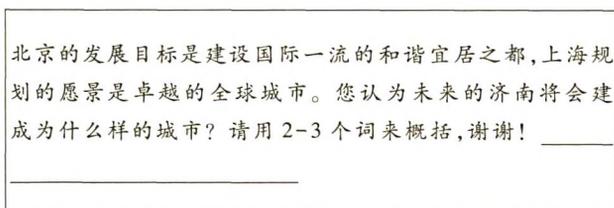


图 2 问卷问题示例

资料来源:作者自绘

### 3.1 愿景文本采集

此次文本采集通过两个方式完成。一个是手工录入,针对的是通过各级政府机构发放到街道(乡镇)一级填写的纸质问卷,发放范围基本涵盖了济南市区全部街道(乡镇)。另一个采集方式是管理后台直接调取,针对的是通过专业调查网站——“问卷星”平台发放的电子问卷。此次城市发展愿景文本数据共计发放问卷 10598 份,回收问卷 8057 份,其中规范填写城市愿景问题的有 4958 条,占回

收问卷数量的 61.54%。采集的文本主要是短文本类型,其中以词汇型为主,也有部分短句型文本。

### 3.2 愿景文本的预处理

预处理首先是进行格式处理。两种形式收集到的问卷数据都需要存储成“txt”格式。同时,在分词处理的环节,本次分析的精度是识别到 2-3 个字组成的基本词汇,所以在 ROST-CM 软件设置中对 4 个字以上组成的词汇进行了过滤,这就能保证把最基本的词汇识别出来进行相关的分析。另外,本环节结合市民填写的具体内容,通过人工添加的形式对 ROST-CM 软件中的词库进行丰富和补充,从而使词汇分离更加精确。补充前后的分词结果,如表 3 所示。

### 3.3 愿景文本的分析及呈现

#### 3.3.1 词频分析

将分词后的 txt 格式文本数据导入到 ROST-CM 软件内,进行词频分析操作,就可以将出现频次较多的单词统计出来,图 3 显示了排名前 20 的单词。在此基础上还可选择词频最多的若干词组进行词云图制作,图 4 是选择了前 300 位的单词,结合济南市的行政区划边界,在 BLUE-MC 词云工具中生成的词云图<sup>[10]</sup>。

表 3 软件词库补充前后分词处理结果差别示例

示例		愿景填写内容
1	处理前	建设和谐,美丽的济南
	处理后	建设 和谐,美丽 的 济南
2	处理前	区域性制造业,物流,文化中心
	处理后	区域 性 制造 业,物 流,文 化 中 心
3	处理前	宜居文化名城
	处理后	宜居 文 化 名 城
4	处理前	交通方便,休闲生活配套完善
	处理后	交通 方便,休 闲 生 活 配 套 完 善
5	处理前	智能,生态和谐文明而文明,安全而舒适
	处理后	智能,生态 和 谐 文 明 而 文 明,安 全 而 舒 适
6	处理前	安全文明,和谐宜居
	处理后	安全 文 明,和 谐 宜 居

资料来源:在参考文献[7]的基础上整理而来

在词频统计结果中,“宜居”是出现频次最高的单词,频次达到 1662 次,其次是“和谐”、“生态”,频次也在 500 次以上,接下来是“文化”、“文明”、“美丽”、“环境”、“泉城”、“创新”和“现代化”,分列 4-10 名,出现频次在 250 次以上,“安全”、“交通”、“人文”和“优美”四个单词紧随其后,出现频次也在 200 次以上。在这里需要说明的是,词频统计只能

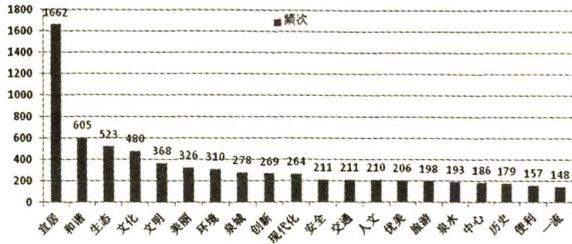


图3 词频统计前20名词汇列表

资料来源:在参考文献[7]的基础上加工而来



图4 词频统计排名前300的词汇生成的济南词云图

资料来源:在参考文献[10]的基础上加工而来

显示出一个静态的分布结果,而无法体现出词汇之间的联系和词群分布情况,而语义网络分析则能满足这一方面的需求。

### 3.3.2 语义网络分析

在 ROST-CM 软件中点击社会网络和语义网络分析按钮,经过运算得到语义分析网络图(图5)。通过计算高频词的中心度和关联度,能将愿景分析的结果进一步细化。

分析高频词的中心度可以发现,中心度计算排名前十名的是宜居、和谐、生态、文化、环境、优美、交通、美丽、创新、文明。这些词汇与其他词汇联系较多,从中也可以以它们为核心进行词群的分类。“宜居”这个词汇是高频词的中心度最高,图中所有的高频词都是围绕着它分布的。这说明宜居不仅

是词频统计次数最高,也是与其它词汇组合频率最高的词,所以城市愿景文本的第一级的核心词汇是“宜居”,这是第一个市民关注的领域。第二级核心词,是除去形容词后,脱颖而出的是2-10名中心度高的词汇中的“生态、文化、环境、交通、创新”。它们分别代表了生态环境、历史文化、交通设施、产业创新等四个领域,由此本次文本分析得到市民愿景表达集中的领域。同时,“中心”和“泉城”两个高频词中心度也比较突出,体现了市民在济南区域地位和地域特色上的愿景诉求。另外,在单词识别的基础上,可以进一步根据词与词之间关联度的数值,对两个字的单词词汇再进行拼接组合,从而形成更高一级四个字的词汇,这样会对市民表达的意愿和需求的程度有更加深刻的认识。以排位前20名的高频词为主题词,结合语义网络的关联度数值,对与其关联的词汇按照一定原则进行拼接(表4),再经过比对汇总,可以得到以下愿景表达中5个等级的拼接词汇。

第一级是“生态宜居”,“生态”与“宜居”二词的词频和关联度均是最高,表达了市民对城市愿景的最高需求。第二级是“文明和谐、交通便利、环境优美”,这三个拼接词的关联度处在100-200之间,体现了市民对城市愿景的次高需求。第三级是“历史文化名城、美丽泉城”,这两个拼接词关联度处在50-100之间,体现了市民对城市文化和特色愿景的表达。第四级是“科技创新、旅游休闲、国家中心、全国一流”,这四个拼接词关联度处在20-50之间,体现了市民对产业和区域定位的愿景表达。另外,还有两个单词,“现代化和安全”,虽然拼接其他单词的关联度并不是很高,但是在愿景的表达中也不应忽视。

表4 主题词及其关联搭配强度统计表

等级	拼接词汇(关联度)	备注
一	生态宜居(298)	200<关联度
二	环境优美(165)、文明和谐(121)、交通便利(110)	100<关联度<200
三	历史文化名城(83)、美丽泉城(57)	50<关联度<100
四	国家中心(46)、旅游休闲(39)、科技创新(28)、全国一流(20)	20<关联度<50

注:词汇关联度的拼接遵循由高到低的顺序,排名靠前的主题词优先与关联词按照关联度进行拼接,排名靠后的主题词选择的关联词不能与前面的重复。

资料来源:在参考文献[7]的基础上加工而来

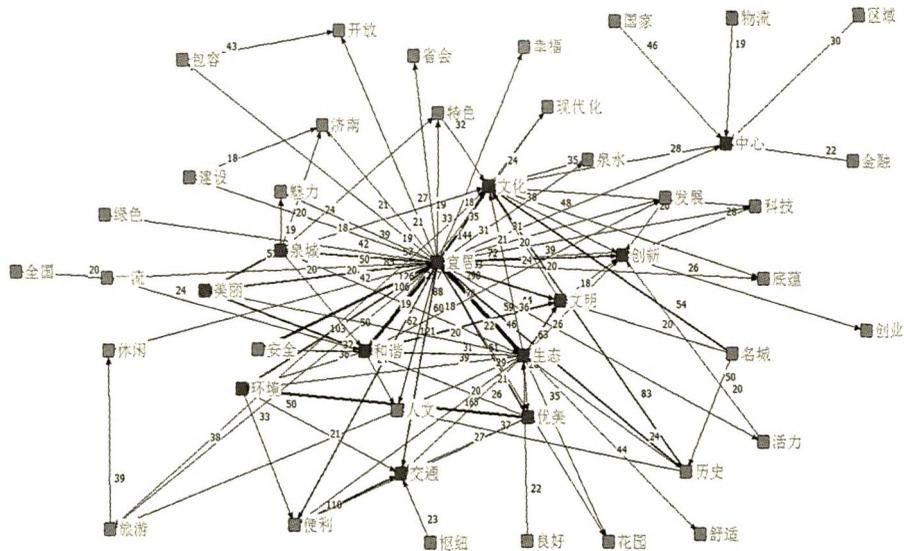


图5 城市愿景填写答案语义网络分析图

资料来源:在参考文献[7]的基础上加工而来

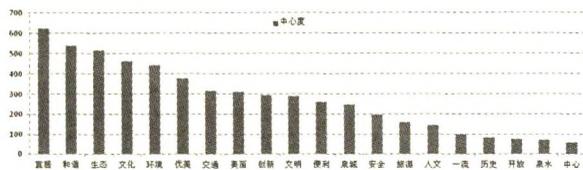


图6 中心度较高的词汇前20名统计图

资料来源:在参考文献[7]的基础上加工而来

表5 词频统计排名前10关键词验证结果

关键词	软件识别数	人工识别数	整体有效率
宜居	1662	1676	99.16%
和谐	605	606	99.83%
生态	523	531	98.49%
文化	480	545	88.07%
文明	368	369	99.73%
美丽	326	326	100.00%
环境	310	312	99.36%
泉城	278	278	100.00%
创新	269	282	95.39%
现代化	264	264	100.00%
合计	5085	5189	98.00%

注:整体有效率=(软件识别数/人工识别数)%。

资料来源:在参考文献[7]的基础上整理而来

### 3.4 分析结果的验证

为了提高分析结果的可靠性,本研究以词频频次排名前10的关键词为例,对软件分析结果进行人工复检验证。通过人工识别发现,软件对词汇的整体识别正确率在98%左右,其中“现代化”和“美丽”两个词的识别率是100%,但是“文化”一词的软件识别率偏低,仅为88%。这是由于市民在“文化”一词

的表达中较为口语化、用词不规范造成的。同时,研究根据文本填写的实际内容对软件中的词库进行补充,但是在补充几轮词库后,我们发现始终有几组词汇其有效率仍然无法达到100%。不过就整体分析而言,本次分析结果仍具有较高的可信度。

## 4 结语

通过对济南城市发展战略规划中愿景问卷填写的文本分析可以发现,济南市民对未来城市“宜居”发展的需求最期待,其关注集中在生态环境、文化特色、交通设施、产业创新等领域,并且在区域定位、现代化和安全方面也有一定诉求体现。虽然完成了相应的规划分析要求,但文本分析是一项较为复杂的分析方法,本研究只是对其在城市规划的公众参与中进行了初步的探索。同时,鉴于中国语言表达的丰富性,在实际的分析中,如果脱离了语境而只是单纯地去分析词汇则会产生歧义,所以也有部分学者认为文本分析方法今天并不适用。但是文本分析的目的主要是在海量的文本中把方向性的观点识别出来,因此我们不仅要认识到它目前存在的精确性不够、对分析者经验依赖性强等问题,更要意识到它分析的快速性和呈现的直观性优点<sup>[7]</sup>,在未来的工作中对其继续优化改善,进而成为城市规划中较为成熟的技术方法。△

致谢:本研究在文本采集过程中得到了原济南市规划局相关工作人员的帮助,在此表示感谢!

## 【注释】

- ①目前界定大数据的标准主要是收集方式和获取数量。如果收集的方式是批量式的,特别是借助相关软件进行网络爬取,同时获取的数据量条数较多,则其数据可以称之为大数据。反之,如果数据获取的方式以手工录入为主,且数据条数较小,则可以称之为小数据。海量数据的收集方式和获取数量介于二者之间。
- ②中心度反映的是词汇在由若干词汇组成的语义网络中的中心性。其计算方式是以该词汇为中心,将与该词有联系的词汇关联度数值加起来求和。中心度越高说明该词在文本中的连接作用越突出,可以据此进行词群的分析。
- ③关联度描述文本中两两词汇之间的联系程度,通常是两个词之间前后相邻出现的次数,这个数值直接能显示在语义网络分析图的连线上。两个词之间的关联度越高,说明二者之间的联系越紧密。
- ④ROST-CM 软件是武汉大学 ROST 虚拟学习团队开发设计的内容挖掘系统,能够对文本数据进行分词、词频统计和语义网络分析等操作。
- ⑤BLUE-MC 平台是由北京蓝色光标品牌管理顾问股份有限公司研发的平台,其中词云工具是该平台的主要产品之一,能够将文本数据生成可视化的词云图。

## 【参考文献】

- [1] 秦萧,甄峰.论多源大数据与城市总体规划编制问题[J].城市与区域规划研究,2017(4):138-157.
- [2] 张磊,王晨.基于内容分析法的中美城市规划公共政策议题比较研究[J].城市发展研究,2011(11):31-38.
- [3] 张峰,陶玉国.游客对于旱景区气候感知的情感分析——

以 5A 景区为例[J].资源开发与市场,2019(8):1093-1099

- [4] 文宏.中国政府推进基本公共服务的注意力测量——基于中央政府工作报告(1954—2013)的文本分析[J].吉林大学社会科学学报,2014,54(2):20-26+171.
- [5] 邱率城.作为现代化手段的城镇化之路——基于四代领导人经典文献的文本分析[C].决策论坛——管理决策模式应用与分析学术研讨会,2016.
- [6] 范珈瑜.基于文本挖掘的游客对古镇旅游态度的分析[J].大数据,2017(06):93-101.
- [7] 李飞.文本分析在规划公众参与中的应用研究——以济南城市发展战略规划为例[D].山东:山东建筑大学,2019.
- [8] 郭安宇.我国城市规划公众参与研究——以红光组团控规修编为例[D].成都:西南交通大学,2016.
- [9] 罗家德.社会网络分析讲义[M].北京:社会科学文献出版社,2010.
- [10] 济南市自然资源和规划局.新时代、新泉城、新梦想——《济南城市发展战略规划(2018-2050年)》解读之一[EB/OL].  
[http://jnup.jinan.gov.cn/art/2019/4/2/art\\_10235\\_2884037.html/2019-04-02](http://jnup.jinan.gov.cn/art/2019/4/2/art_10235_2884037.html/2019-04-02)

作者简介:赵虎(1982-),男,山东茌平人,博士,山东建筑大学建筑城规学院副教授,硕士,研究方向为城市与区域规划。

收稿日期:2019-08-20

## Application Research of Text Analysis in Planning Public Participation: A Case Study on Analysis of Urban Vision in "Jinan City Development Strategic Planning"

ZHAO Hu, LI Fei, CHEN Yu

**【Abstract】**With the upgrading of network information technology and the popularization of Handheld terminal equipment, Citizens can more easily participate in urban planning, and generate large amounts of text data, so there is a need for matching analysis methods to adapt to this. This study defines the types and needs of public participation in Text Analysis firstly. Then proposed a five-step technical framework for large-scale text analysis. Finally, it applies the case of the city vision of Jinan Urban Development Strategic Planning and gives corresponding results. Through the use of ROST-CM software for word frequency analysis and semantic network analysis, in Jinan urban development strategy planning, the primary demand of citizens for urban vision is "livable", and the demand areas are concentrated in ecological environment, history and culture, transportation service, industrial innovation, regional positioning and local characteristics, then gives different levels of vision requirements.

**【Keywords】**Text Analysis; Urban Planning; Public Participation; City Vision; Jinan City