

基于数字政府建设的公文标签体系构建与应用研究

金加和¹, 施筱玲¹, 徐峰², 叶红叶¹, 俞巍滔¹

1. 浙江省数据开放融合关键技术研究重点实验室, 浙江 杭州 310007;

2. 浙江省高层次人才服务中心, 浙江 杭州 310007

摘要

公文作为政府机关办公、办事、办会的主要载体, 在数字政府建设中起着信息记录、交换、传递和指令发布等作用, 构建公文标签体系是促进政府机关公文高质量办理和提升政府行政效率的重要途径。针对公文标签, 提出了基于数字政府建设的公文标签体系构建方法, 开发设计了以业务场景驱动与用户需求为导向为核心的政府公文标签体系。该体系通过公文标签的设计、生成与管理创新, 旨在破解政府公文管理效率低下、内容分散、流程烦琐等难题, 实现政府公文管理与运行效率的提升。同时, 对公文标签在不同场景下的创新应用进行了阐述, 并给出了基于政府机关的公文标签应用案例, 为推进数字政府高质量发展提供重要支撑。

关键词

数字政府; 公文标签; 应用场景; 智慧政务

中图分类号: TP399

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024061

Research on the construction and application of official document label system based on the construction of digital government

JIN Jiahe¹, SHI Xiaoling¹, XU Feng², YE Hongye¹, YU Weitao¹

1. Zhejiang Provincial Laboratory of Data Integration Technology, Hangzhou 310007, China

2. Zhejiang Provincial High Level Talent Service Center, Hangzhou 310007, China

Abstract

As the main carrier for government agencies to handle documents, services, and meetings, official documents play the role of information recording, exchange, transmission and instruction release in the construction of digital government. The establishment of an official document labeling system is an important means to promote the high-quality handling of government documents and improve the administrative efficiency of the government. Aiming at the official document

label, a method for constructing an official document labeling system based on digital government construction was proposed, and a government official document labeling system with the core of business scenario drive and user demand orientation was developed and designed. Through the design, generation and management innovation of government official document labels, this system aims to solve the problems of low efficiency, scattered content and cumbersome processes of government official document management, so as to improve the efficiency of government official document management and operation. At the same time, the innovative application of official document labels in different scenarios was expounded, and the application cases of official document labels based on government agencies was given, which provides important support for promoting the high-quality development of digital government.

Key words

digital government, official document labeling, application scenario, smart government affair

0 引言

随着数字化时代的到来,数字政府建设已经成为各国政府提升行政效率的重要任务。数字政府建设旨在利用信息技术和数字化手段,建立健全、标准化、规范化的制度规则,促进政府行政管理的数字化和智能化,推动政府治理体系和治理能力的现代化。公文是数字政府机关办文、办事、办会过程中形成的文本文件、演示文档、视频资料等文本信息,是数字政府机关日常工作中不可或缺的一部分,它承载政府的决策、指示、执行和服务等重要信息,对数字政府运行和公共事务的推进发挥了至关重要的作用。

从研究者角度来看,公文是法定的社会组织在处理公务过程中形成并使用,具有法定效力和规范体式的文书^[1],也是政府各部门内部管理过程中产生数据中的一种^[2]。传统意义上,政府管理基本上单纯依靠正规渠道的随机调查统计数据进行决策,但这种决策方式在当今大数据时代背景下无法持续被使用^[3]。在数字化改革持续深化、政府数据价值日益凸显的背景,政府公文管理也面临一系列的挑战。一方面,公文数据资源已经不再局限于传统

的信息检索和报表分析,需要支持业务状况实现智能感知、实时分析,需要通过对跨领域、跨部门、跨层级、跨系统的数据建立主题联接,支持对业务的关联影响分析以及对目标对象的特征识别。另一方面,目前公文管理方式存在效率低下、信息不透明、流程烦琐等问题,而且公文内容分散、缺乏语义关联,公文之间未形成有效关联^[4],缺乏体系化标签体系,限制了公文间的相互引用,影响公文内容的一致性。公文管理场景需求的个性化、多元化与政府数据资源的海量无序之间的矛盾日益突出,已无法满足数字化时代对公文管理的要求。因此,对公文进行分类打标签,用于公文检索、推荐和分析等智能化管理,以提高政府办文办事效率,当前尤为迫切。标签是根据业务场景的需求,通过对具有静态、动态特性的目标对象运用抽象、归纳、推理等算法得到的高度凝练的特征标识,以便于被检索和共享。一般而言,标签管理应包括标签体系建设、标签库建设、标签模型训练、标签应用,其主要作用是通过特征集合并关联被打标签的对象,将分析对象生成画像,挖掘对象的价值,赋能数字化应用场景。

公文标签作为对法定机关或其他社会组织处理公务往来的文件、会议、活动等进行业务实体特征描述的数据形式,是推动数字政府公文处理规范高效的基础工

具。利用公文标签,可对跨领域、跨部门、跨层级、跨系统的数据建立主题连接,能有效满足智能检索、智能推荐、智能分析等场景应用的需求。公文标签已成为数字政府建设中必不可少的重要内容,一套科学规范的公文标签体系亟须被构建并加快应用。在实际应用中,公文标签体系可以与政府的办文、办会、办事等业务信息系统进行对接集成,实现公文运转的自动化、高效化。将公文标签与政府业务信息系统相结合,可以实现公文的自动归档、自动检索和自动分发,提高公文管理的效率,增强公文管理的便利性。

本文基于政府机关的业务需求,研究面向大规模政府公文数据资源的标签体系的构建,并分析公文标签体系的应用场景,对提升政府公文管理标准化、数字化、智能化水平,促进数字政府高质量发展具有重要意义。

1 研究综述

为充分挖掘公文数据资源的价值,提高政府公文处理的工作效率,近年来,我国出台了相关政策和标准,尝试运用标准化、数字化手段加强政府公文管理。2016年12月30日,中华人民共和国国家质量监督检验检疫总局、中国国家标准化管理委员会发布《党政机关电子公文元数据规范》(GB/T 33480-2016)和《党政机关电子公文标识规范》(GB/T 3347-2016),对电子公文的核心元数据、标识编码规则进行了规范。2020年11月19日,国家市场监督管理总局、国家标准化管理委员会发布《党政机关电子公文归档规范》(GB/T 39362-2020),对党政机关电子公文归档流程、格式等做了明确要求。2004年4月5日,中华人民共和国国家质量

监督检验检疫总局、中国国家标准化管理委员会发布《电子政务主题词表编制规则》(GB/T 19486-2004),明确了电子政务主题词表的编制原则、方法。在地方层面,已有福建、天津、湖南、广西等省市发布并实施电子公文的基础数据元和数据存储、数据交换以及文件归档等技术规范。总体来看,在国家和省级层面,目前公文电子化规范建设基本上是对电子公文的流转流程、标识规范、档案编码等进行流程性规范,对与公文主题相关的数据标签规范建设处于“空白”状态。

在学术研究领域,学者对公文智能化处理的研究主要集中在4个方面:公文智能推荐、公文智能分类、公文辅助写作以及公文智能审查。

(1) 公文智能推荐

张真继等^[5]在电子政务系统中研究了“主题智能公文包”服务,提出在系统内建立利于用户自身习惯的方便快捷的信息定制服务(即主题智能公文包)功能模型;皎海军等^[6]则研究了基于词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)算法的公文处理智能推荐技术,通过TF-IDF算法挖掘历史公文,构建用户画像。

(2) 公文智能分类

规范的公文数据标签体系是实现公文智能化处理的前提,现有的对公文数据标签体系研究的文献内容主要集中在文本分类技术,包括单标签分类、多标签分类以及网络化标签分类三大类。其中,单标签分类不考虑关联性;多标签分类考虑了关联性;网络化标签分类则通过对内在关联性进行建模,挖掘多标签的相关性^[7]。在具体实践研究中,黄永等^[8-9]利用词汇特征和机器学习方法对学术文本的功能结构进行自动分类,Gasparetto等^[10]进行了文本分类算法的调查。

(3) 公文辅助写作

马嘉荣等^[11]以WPS写作为例,分析了智能化工具对公文写作的积极影响,提出智能化工具会促进公文格式的规范化、提高党政机关的行政效率、推动公文写作知识的整合、降低公文写作难度。柏峰等^[12]设计开发了公文智能辅助写作系统,在机关单位人员撰写公文时,可检索、推荐相似的政府公文作为参考。

(4) 公文智能审查

近年来,深度学习因具有强大的特征选择和特征抽取能力,能够自动捕获文本更高层次、更抽象和更全面的语义信息,在政策文本研究中得到了应用^[13]。陈思^[14]提出将深度学习技术应用于电子公文智能审查,基于深度学习的算法进行研究,具体包括针对电子公文的公文印章、正文信息提取等应用。赵洪等^[14]提出在数量庞大的各类公文资源中,亟须依托大数据分析 with 人工智能技术,采用科学化、智能化处理方法,实现公文的科学流转、精准匹配和高效办理。构建公文标签体系是加快推动政府数据治理、实现数据高质量供给的一种有效手段,它对政府数据的汇聚关联、质量提升、开发利用和价值实现具有重要作用。杨孟辉等^[3]提出了政府大数据治理的观点:一方面是管理政府大数据;另一方面是利用政府大数据,认为政府数据的唯一性、高质量和高价值密度是政府决策和向公众提供公共产品和服务的重要支撑。政府数据治理要着眼于政府部门之间的内部数据流动关系,强调数据处理过程中的强关联性,主张数据治理从分散走向集中、从部分走向整体、从碎片走向整合^[15]。

综上所述,当前我国政府公文管理的理论研究明显滞后于实践需求,公文标签作为数字政府建设的一项重要基础内容,对其应用研究明显不足,数字政府建设缺乏有力支撑,亟须对公文标签体系建立一

套新型研究方法,旨在满足数字政府日益发展的公文处理多样化、智能化的需求,同时,为相关技术研究和政务系统开发建设提供可供研究、可借鉴的思路与方法。

2 公文标签体系设计流程

政府公文是一种重要的数据资源,随着政务数据资源的应用需求日益增多,迫切需要编制统一的公文数据标签技术规范,指导数字政府领域各部门、单位等基于统一的数据标签规则进行公文数据管理。公文标签体系的构建是指将公文按照一定的标准和规范进行分类和标识,形成一套通用的、易于理解的标签体系,以方便公文的检索、查阅和管理。公文标签体系是一项系统性的工程,在设计之初就需要规划好流程、原则等,本文提出如下公文标签体系设计理念、设计原则、生成方法及设计流程。

2.1 公文标签体系设计理念

2.1.1 业务场景驱动

以业务场景倒推标签需求,避免标签框架大而全,尽量使标签与业务场景贴得更紧密。不同用户对政策标签的诉求有较大的差异,不同业务场景关注的标签也存在较大差异。大而全的标签框架实际是从顶层用户视角搭建的,对具体业务场景的适用性较差,因此应该从业务视角来实现,放弃顶层的用户抽象视角,针对各业务线或部门的诉求和实际的应用场景,分别将标签聚类起来提供给相应部门。如社保部门更关注与社保相关的标签,财政部门则更关注与财政税收相关的标签,教育部门更关注与教育相关的标签,因此不可能用一套标签覆盖整个业务团队。这种从业

务场景倒推标签需求的方法,能够使标签与业务场景贴合得更紧密,使可用性提升。

2.1.2 用户需求导向

从服务用户的视角,围绕业务的时代性、重要性、利益性等特征来设计。如政策文件库惠企主题分类包括减税降费、复工复产、创业扶持、产业发展、财政支持、科技创新、转型升级、人才政策、融资服务、乡村振兴,主要服务用户为中小企业,是参考了主流中小企业政策服务平台的主题分类以及有关疫情下中小企业对政策刚需的调研结果提出的。

标签层级尽量不超过二级,建议附注解。因为参与标签体系构建的人员可能并不是直接进行数据标注的人,且不同业务线对标签定义的理解不同,如果不将标签的具体含义以及相应案例附上,很容易造成实际标注效果不尽如人意的结果。

2.2 公文标签体系设计原则

在对政府公文设计标签时,应遵循以下原则。

(1) 权威性

分类体系优先遵循标准。如发布机构可参照中国政府网的国务院组织机构设置,行政区划可参照《2020年中华人民共和国县级以上行政区划代码》,企业类型可参照《关于划分企业登记注册类型的规定》,企业规模可参照《统计上大中小微型企业划分方法》,行业分类可参照《国民经济行业分类》(GB/T 4754-2017)等。

(2) 先进性

体现跨域、智能要求,通过标签体系规范化、标准化建设,促进公文数据资源跨部门、跨领域、跨层次增值化利用,可提供智能搜索、智能问答、智能推荐、智

能分析等服务。同时,利用机器学习和深度学习等技术对政府文件进行智能化分析和处理,提高标签匹配度和命中效率。支持与AIGC(AI-generated content)大模型和多模态知识图谱(multi-modal knowledge graph, MMKG)等新技术进行无缝对接,为推进政务大模型建设提供基础性、高质量服务。

(3) 科学性

分类体系应按照“相互独立,完全穷尽”(mutually exclusive collectively exhaustive, MECE)的原则进行设计制定,做到不重不漏。

(4) 好用性

结构清晰,每一个子集的组合都能覆盖父集所有数据,并具有关联性。分类层级尽量不超过4级。如标签“乡村振兴”的分类层级,见表1。

通过数字政府公文处理系统搜索“乡村振兴”,可以搜索出《国家发展改革委关于恢复和扩大消费措施的通知》,但该文件中没有乡村振兴相关表述,按传统关键词搜索,无法通过“乡村振兴”关键词搜索到该文件。但在设计标签规则时包含了“千万工程”表述,根据标签规则关联了乡村振兴标签,这样即可成功定位与乡村振兴相关的政策文件。

2.3 公文标签体系生成方法

标签生成主要有两种方法:①机器标

表1 标签“乡村振兴”的分类层级

一级标签	二级标签	三级标签	标签规则
经济建设	乡村振兴	乡村振兴综合	农业农村现代化;乡村振兴示范省;“千村示范、万村整治”工程;美丽乡村;美丽城镇;“千万工程”;美丽县城、美丽城镇、美丽乡村联创联建

注方法,即系统自动打标签,通过分析文本、图片等信息,结合规则模型算法的自然语言处理技术,实现高精度的匹配,自动输出内容标签;②人工标注方法,即业务人员根据业务需求及个人理解对政务信息内容进行标签标注。为了确保标签的有效性,需要对标签进行质量评估,并通过质量评估结果不断优化标签体系,从而为标签应用提供有效的标签数据。标签生成的具体流程如图1所示。

此外,文本分类是自然语言处理领域中许多任务的基础。Rocchio^[16]提出Rocchio分类算法,后续的研究更多是基于这种分类方法,并将之延伸到大规模文本分类(Partalas等^[17])和多标签分类(Ali等^[18])上。Hinton等^[19]创造性地提出了深层次的神经网络训练方式,深度学习的方法从此开始发展,在图像和自然语言处理方面都得到了很好的运用,基于深度学习的文本分类方式也成了主流。这些模型大大提升了文本分类和标注的效率。

利用基于规则和基于深度学习的模型分析各维度的数据特征,进行机器自动打标,节约了人工判断成本。基于规则和深度学习的特征分析模型进行标签生成的流程如图2所示。

2.4 公文标签体系设计流程

标签体系的构建主要包括业务梳理、标签分类、标签定义和标签管理四大步骤,最终实现标签的场景应用。

2.4.1 业务梳理

业务梳理可极大程度上保证标签的全面性与合理性,因为“以业务为导向”的标签才具有真正的生命力,真正地做到反映客观实际。如果想推动某个惠企政策的落实,需要创建这个政策的主题关键词,同时需要创建对应符合该政策的各项标签,如“所属行业”“企业规模”“申请历史”

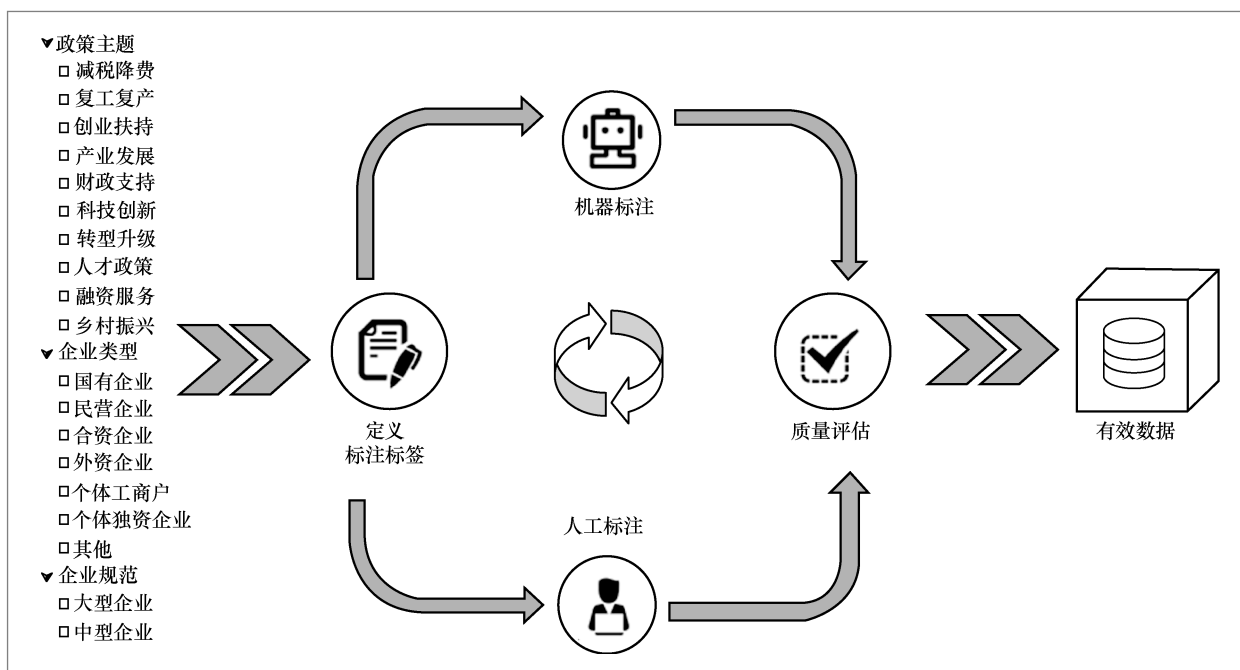


图1 标签生成的具体流程

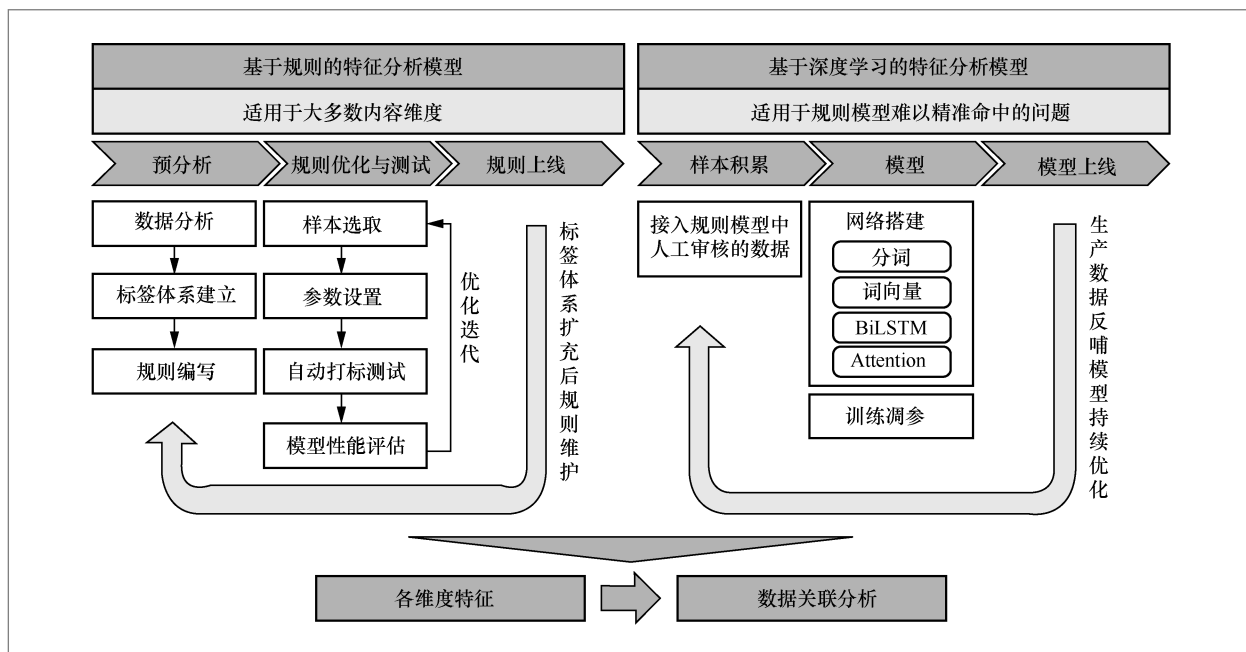


图2 基于规则和深度学习的特征分析模型进行标签生成的流程

等，还需要考虑该政策从哪些渠道传达给符合要求的企业，让其能及时、准确地获知该政策的相关信息。

可以按以下思路来梳理标签体系：

①梳理有哪些业务线，每个业务线有哪些来源渠道，将其一一列出；②每条业务线有哪些服务对象和资源，比如用户、政策、留言等；③根据服务对象聚合业务，每个对象涉及哪些业务场景，每个业务场景下涉及哪些业务数据和用户行为。

梳理业务场景时，可以先对场景进行分类，分为主动推送用户和用户行为被动触发两大类；然后在两个分类中进行梳理。其中每个场景可以由以下几部分组成。

- 场景名称：用于简单描述和标示场景。
- 触发机制：可以是单次手动触发，也可以是周期性循环或者某行为时间触发，用于标示本次场景开始的时间节点。
- 用户范围：用于标示本次场景涵盖的用户范围，建议划分为用户属性和用户行为。
- 内容范围：用于标示本次场景提供

的内容范围，可以是政策，也可以是服务，因此又可分为政策标签与服务标签。

- 触达方式：选择用户触达政务信息服务的渠道，可以是网站、小程序或者App等。

业务梳理示例如图3所示。

2.4.2 标签分类

按业务需求梳理业务数据后，可以继续按照业务产出数据的属性进行分类与标签设计。标签本身会有很多分类，但是按照数据生成和处理的维度，标签可分为事实标签、规则标签、预测标签，这3类标签的生成难度依次递增。这种标签的分类方式更多是面向技术人员，帮助他们设计合理的数据处理单元。

(1) 事实标签

既定事实，从原始数据中提取。事实标签反映事实的基础特征，对数据的加工程度最低，主要通过文本挖掘的方法，从原始数据中尽可能多地提取事实数据信

<p>场景1: 为中小企业用户提供惠企税收优惠政策服务</p> <p>触发机制: 用户主页访问</p> <p>用户标签: 企业类型为中小微企业</p> <p>政策标签: 政策主题分类为税收优惠, 政策企业分类为中型企业、小型企业、微型企业</p> <p>触达方式: 用户主页的“政策推送”</p> <p>(a) 为中小企业用户提供惠企税收优惠政策服务</p>
<p>场景2: 为应届毕业生用户提供自主创业优惠政策服务</p> <p>触发机制: 用户主页访问</p> <p>用户标签: 应届毕业生 (基本信息中获取)</p> <p>服务标签: 办事服务对象为应届毕业生, 且服务类别为创业优惠的服务事项</p> <p>触达方式: 用户主页的“我的服务”</p> <p>(b) 为应届毕业生用户提供自主创业优惠政策服务</p>

图3 业务梳理示例

息, 如政策发布单位、发文时间、发文类型、发文平台等。

(2) 规则标签

规则标签拥有很强的业务属性, 需要业务人员和数据分析师一起分析探索, 基于原始数据进行标签规则的创建和拼接。比如要创建一个“应届毕业生”标签, 需要从各个维度对“应届毕业生”进行特征定义, 比如年龄方面一般为20~25岁、是否办理过就业补助、是否访问过人才引进政策信息等。

(3) 预测标签

预测标签主要根据业务需求定义预测的场景, 建立预测模型, 结合实际的业务数据形成预测标签。例如根据某类政策访问用户的共性特征来预测其他用户对此类政策的偏好程度。又如根据流失用户群的特征来预测其他用户群的流失概率。预测

标签需要基于过往数据和机器学习算法不断地进行模型优化调整来提升预测的准确度。

2.4.3 标签定义

标签定义主要包括标签实体定义、标签属性定义、标签主题定义。通过标签定义主要说明标签是什么、标签来源于哪里、标签如何计算、标签何时生成等。

- 标签实体定义。实现标签主体的定义, 包括实体名称、数据来源、相关数据集、同步计划等。

- 标签属性定义。实现某个实体下具体标签的定义, 包含标签名称、标签代码、标签类型、生成规则、值域等。

- 标签主题定义。根据实际用户分析需求, 在标签定义的基础上, 提供快速定义分析主题的功能, 满足用户分场景的查询需求。

标签定义的部分示例见表2。

2.4.4 标签管理

标签管理与维护往往是容易被忽视的点。事实上, 标签也具有生命周期, 包括从需求提出到生成、审批、执行的整个流程。对于很多政府单位来说, 生成一期标签并不难, 目前已有部分政府机构梳理了三四百或者上千的标签, 但是大部分标签在被生成之后, 并没有明确的更新规则、标签更新维度、标签更新权限、无用标签的淘汰, 如标签库内只会使用60个标签, 但是标签库中有90个标签, 其中有很多无效标签占用资源等。因此, 标签的更新迭代和维护管理是一项非常重要的工作。

在标签体系设计后, 在具体应用前需要对标签进行管理。标签的管理与维护可

表 2 标签定义的部分示例

一级标签	二级标签	三级标签	四级标签 (标签实例)	规则定义	标签类型
人口属性	基本信息	性别	性别-男	注册信息获取	事实标签
			性别-女	注册信息获取	事实标签
			性别-未知	注册信息获取	事实标签
		年龄	年龄-××岁	注册信息获取	事实标签
			生日	生日-××××/××/××	注册信息获取
行为属性	上网习惯	终端类型	终端类型-PC	系统标注	事实标签
			终端类型-移动	系统标注	事实标签
	活跃情况	活跃情况-核心用户	满足其中条件之一即为核心用户： 1.过去30天内，发生a行为3次及以上； 2.过去60天内，发生b行为6次以上。	规则标签	
		活跃情况-活跃用户	满足其中条件之一即为核心用户： 1.过去30天内，发生a行为1次及以上； 2.过去60天内，发生b行为2次以上。	规则标签	
	业务属性	内容偏好	招标采购信息	模型预测	预测标签
			涉农补贴信息	模型预测	预测标签

通过搭建标签管理系统来实现，如完成标签库创建、数据标注、模型训练与管理、标签体系迭代优化，并对标签体系进行持续运营管理，确保标签体系的可靠性。标签管理流程如图4所示。

通过标签体系的持续运营管理，可以查看模型的运行状态，还可以查看模型对线上数据的处理情况，详细查看模型的结果是否正确，并在对不正确的结果进行人工修正后，可将这些数据添加到训练数据中，不断增加训练数据，用于模型

的优化完善，以及政策图谱的设计开发和应用。

3 政府公文标签体系构建与应用

为有效挖掘大规模政府公文数据资源，支持政府公文流转工作中的智能公文管理和智能化决策，本文利用机器标注和人工标注方法，坚持权威、先进、科学、好用性原则，以推动数字政府建设和提高政

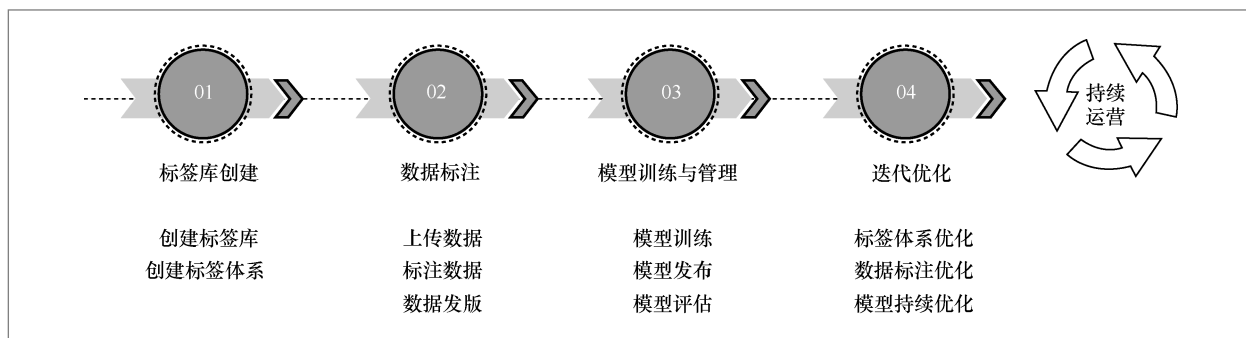


图 4 标签管理流程

府政务智能处理能力为目的,构建了一套4层政府公文标签体系。

3.1 4层政府公标签体系的设计

公文标签体系架构可以分为4层:基础数据层、数据处理层、数据业务层、数据应用层,如图5所示。每个层级面向的用户对象不一样,处理的事务有所不同。层级越往下,与业务的耦合度就越小;层级越往上,业务关联性就越强。

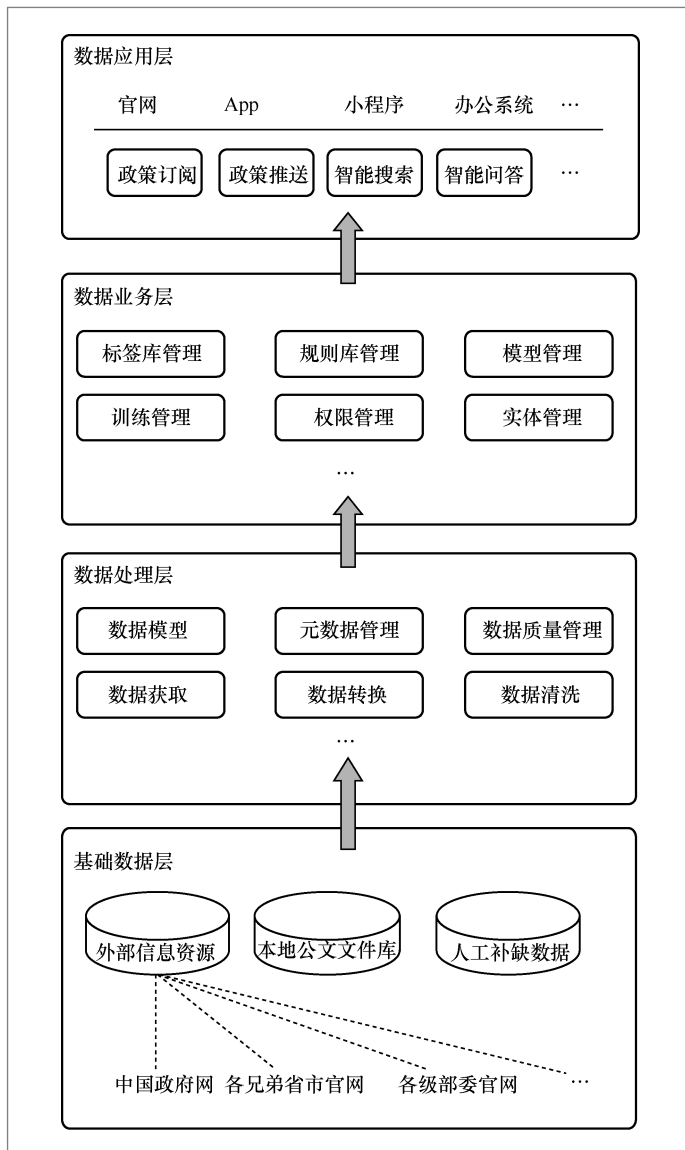


图5 公文标签体系架构

(1) 基础数据层

基础数据层指为数字政府提供支持 and 基础保障的数据架构。这一层的主要目标是收集、存储、管理和分析政府机构产生的大量数据,用于支持政府决策和提供公共服务。在构建数字政府的基础数据层时,需要建立一个统一的数据收集和存储系统。这个系统应该能够从各个政府机构和部门收集数据,并将其存储在一个集中的数据库中,确保数据的一致性。

(2) 数据处理层

数字政府建设的主要目标是确保数据的准确性、完整性和安全性,并提供高效的数据访问和处理功能。在构建数字政府的数据处理层时,首先需要设计建设一个强大的数据存储和管理系统。这可以包括选择合适的数据库技术和架构,以及定义数据模型和表结构。此外,还需要考虑数据的备份和恢复策略,以确保数据的可靠性和可用性。其次,数据处理层需要实现数据的采集和清洗功能。这包括从各种来源收集数据,并对其进行验证、清理和转换,以确保数据的一致性和准确性。同时,还需要实施数据质量控制措施,例如数据去重、异常检测检测和纠错等,用于提高数据的质量和可信度。

(3) 数据业务层

数据业务层指在数字政府中负责数据管理、数据分析和数据应用的部门或团队。它承担着收集、整理、存储和分析政府数据的重要职责,为政府决策提供科学依据和支持。在构建数字政府的数据业务层时,需要考虑建立一个统一的数据资源管理平台,用于集中管理政府各部门的数据资源。这个平台应具备数据采集、数据存储、数据清洗和数据共享等功能,确保数据的质量和安全性。

(4) 数据应用层

应用层的任务是赋予标签实际应用的

能力,聚合业务数据,提供数据应用服务。业务方能够根据自己的需求来使用、共享业务标签,但彼此的业务又互不影响。实践中可应用于政策订阅、政策推送、智能搜索、智能分析、智能问答等智慧政务服务场景。

首先,该标签体系应该基于政府内部的业务需求和流程,对公文进行分类和标识,因此需要考虑公文的类型、主题、涉及的部门和政策领域等因素。例如,可以针对不同类型的公文(如通知、函、意见、决定等)设立不同的标签,每个标签又可以细分为不同的子标签,以便更准确地描述公文的内容和性质。其次,该标签体系应该具备智能化和自适应能力,能够根据用户输入的关键词或标签,自动推荐相关的公文和信息。这需要利用数据挖掘、自然语言处理和机器学习等技术对公文的文本内容进行分析和处理,提取出关键词和标签,并建立相应的索引和推荐算法。最后,该标签体系应该具备可扩展性和可维护性,能够适应政府业务需求的变化和公文种类的增加。这就需要采用可扩展、高性能的技术架构和开发方式,将标签体系与其他信息系统进行集成融合和互联互通,实现数据的高效共享和流通利用。

3.2 标签体系应用研究

3.2.1 智慧办公:决策与数据分析场景

智慧政务办公指利用先进的信息技术手段,提升政府办公效率和服务质量的一种新型管理模式。随着数字技术的迅猛发展,智慧政务办公已经成为现代政府建设的重要组成部分。智慧政务办公的核心目标是实现政府工作的数字化、网络化和智能化。通过建立统一的信息平台,政府部门可以实现信息共享和协同办公,提高工作效率和决策服务水平。同时,智慧政务办

公可以提供更加便捷高效的公共服务,满足人民群众的多样化需求。智慧政务办公的实施需要依托先进的数字技术,如云计算、大数据、人工智能等。通过这些技术,政府可以实现数据的快速采集、分析和利用,为政策制定和决策提供科学依据。同时,智慧政务办公还可以提供智能化的办公环境和工具,帮助办事人员提升工作效率。智慧政务办公的推行,一方面需要政府部门的积极参与和支持;另一方面,需要政府加强对智慧政务办公的宣传和培训,提高工作人员的数字化素养和工作技能。

3.2.2 政策/服务推荐场景

基于用户的个人偏好和历史行为数据,智慧政务服务可以推荐相关的政府服务和信息。例如,根据用户的兴趣爱好和地理位置,推荐附近的政府活动、社区服务等。具体应用如下。

- 在信息页中提供“相关阅读”“相关服务”等类似模块作为扩展功能,将关联信息作为已读信息的补充,将具有关联性的主题进行捆绑式推送,可以提高政府信息资源的“易理解性”,如商务部官网。

- 通过信息系统的后台、微信、短信、邮箱或者App推送给用户其感兴趣的信息,变“被动服务”为“主动服务”。如烟台市对国家、省、市三级现行有效的政策文件进行精细化政策梳理,搭建标签化政策库,同时整合归集社保、税收等企业数据以及民政、残联等民生部门特定政策扶持群体的个人数据,为实现政策精准推送和“免申即享”打造专属档案。

3.2.3 政策/服务订阅场景

用户可以根据自己的需求和兴趣,订阅政府机构发布的信息和通知。政府机构

可以根据用户的订阅情况,定期向用户推送相关的政务信息和服务信息。这样,用户就可以及时了解到最新的政策文件和公共服务,政府机构也可以更好地与公众在线沟通交流。

3.2.4 智能搜索与智能互动问答

用户可以通过关键词搜索政府机构发布的信息和服务。搜索引擎可以根据用户的搜索词和搜索历史,为用户提供相关的政务信息和服务,以此增加政务服务的便利性、提高政务服务的效率。智慧政务服务还提供了强大的互动功能,用户可以通过政务网站或移动应用与政府机构进行在线互动交流。例如,用户可以提出问题、反馈意见、申请办事等,政府机构通过智能化的系统,及时回复用户的问题和意见,并提供个性化的办事指导和服务。这样,政府机构就可以更好地了解公众的需求和意见,不断提高政务服务的满意度和质量。

智慧政务服务依托智能推荐、订阅、搜索和互动等场景,为政府机构和公众提供智能化、个性化的政务服务。这些政务服务的广泛应用,不仅提高了政务服务的效率和质量,也增强了政府机构与公众之间的互动和沟通。未来,随着人工智能和大数据技术的不断发展,智慧政务服务将会得到进一步的完善和拓展,服务将更加便捷和高效。

3.3 公文标签应用案例

3.3.1 政府OA应用系统

在某省政府机关的OA系统中对所有文件进行标签化,根据文档内容的相似性和关联性对文档进行自动分类,从而提高文档的分类精度,同时可以根据某个文件

的所属标签,快速查询该标签的所有关联文件。开发智能检索应用功能,支持通过标签对收发文、领导批示、政务信息等进行全量检索,并展示标签关系图及标签与知识的关联。

智能标签技术的应用,一方面提高了文档的分类精度,通过相关性强的、个性化的特征关键词对事物或内容进行描述,系统将自动根据描述规则为文件打上标签,完成对文件的结构化和标签化,快速进行内容分类,实现文件的精准检索和智能推荐。另一方面提高了文档的检索效率,智能标签技术可以根据用户的需求和文档内容自动为文档打上标签,方便用户快速地检索所需的文档,同时可以根据某个文件所属标签快速查询该标签所有相关联的文件。

3.3.2 政务信息智能采编系统

在某省政府机关的政务信息智能采编系统中应用了智能标签技术,提供了便捷检索、辅助选题、关联参考等功能。其中,便捷检索功能可根据公文标签服务对上报信息进行检索,通过调用经济建设、政治建设、宣传思想和文化建设、社会治理和社会建设、生态文明建设等标签精准索引相关信息。将原来的关键字检索扩展为标签属性检索,拓展了检索渠道。辅助选题功能则通过业务模型将上报信息与任务库、政策文件库等进行标签智能匹配,并通过匹配规则将上报信息进行预分类,辅助政务信息选题甄别,提高编刊前的选题效率。关联参考功能可根据公文标签自动关联与本期报送内容有相似业务属性的往期报送信息,减少重复报送。信息部门进行刊物编辑时,系统自动关联与刊物内容有相似业务属性的往期信息,便于用户及时调整编刊方向,减少重复内容。上报单

位报送信息或信息部门进行刊物编辑时,系统可根据公文标签自动关联与报送内容或刊物内容有关的政策文件信息,供用户参考。

该系统的应用取得了显著成效。一是提高了信息报送质量,截至目前,通过公文标签体系给15 701条信息打上标签,自动关联往期上报、相关政策等信息4 710条,对重复性信息可给予预警,累计预警785条。二是提高了信息选题效率,系统通过公文标签对上报信息进行判定,筛选出符合的相关政策、报送要点、报送任务以及是否约稿等方面的信息,并标为重要信息,在选题时可重点关注重要信息,从而节约报送信息的浏览时间,提高选题效率。三是提高了信息编刊质量,编刊人员在编辑刊物时,系统根据公文标签自动推送过往类似期刊、期刊领导批示、相关政策文件等,辅助编刊人员对编刊内容进行考量,实现精准采编、快速成稿。

4 结束语

本文提出了公文标签体系设计理念、设计原则、生成方法以及设计流程,并在此基础上构建了4层标签设计体系,分析了其应用场景,旨在实现更高效、更便捷的公文管理和智能服务。在数字政府建设的大背景下,公文标签体系构建与应用对政府机关的数字化、智能化运行具有重要的现实意义。本研究具有一定的理论价值和实践价值,为政府公文数据资源价值的实现提供了方法。但也存在一些不足,比如实证性和国际化研究,未来将予以深入。可以预见,公文标签体系应用前景广阔,公文标签体系不仅可以用于公文的分类、检索和智能管理,还可以与其他数字化应用系统,如与政府决策支持系统、政务大模型

等进行集成,可支撑发展高效协同的数字政务,为推进政府治理体系和治理能力现代化提供有力支撑。

参考文献:

- [1] 王红霞,冯国强.公文相近概念及文种的异同比较[J].办公室业务,2011(9):22-23.
WANG H X, FENG G Q. Comparison of similarities and differences in similar concepts and types of official documents [J]. Office Operations, 2011(9): 22-23.
- [2] 王芳,陈锋.国家治理进程中的政府大数据开放利用研究[J].中国行政管理,2015(11):6-12.
WANG F, CHEN F. Openness and exploitation of government big data in the course of state governance[J]. Chinese Public Administration, 2015(11): 6-12.
- [3] 杨孟辉,杜小勇.政府大数据治理:政府管理的新形态[J].大数据,2020,6(2):3-18.
YANG M H, DU X Y. Big data governance in governments: a new form of the government administration[J]. Big Data Research, 2020, 6(2): 3-18.
- [4] 赵洪,王芳,王晓宇,等.基于大规模政府公文智能处理的知识发现及应用研究[J].情报学报,2018,37(8):805-812.
ZHAO H, WANG F, WANG X Y, et al. Research on construction and application of a knowledge discovery system based on intelligent processing of large-scale governmental documents[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(8): 805-812.
- [5] 张真继,刘红璐,孙洁.电子政务系统引入“主题智能公文包”服务的研究[J].东北师大学报(自然科学版),2005,37(4):38-42.
ZHANG Z J, LIU H L, SUN J. Study on thematic and intelligent portfolio service in e-government system[J]. Journal of Northeast Normal University (Natural Science Edition), 2005, 37(4): 38-42.
- [6] 皎海军,杜胜贤,廖晨阳,等.基于词频-逆

- 文档频率(TFIDF)算法的公文处理智能推荐技术研究[J]. 办公自动化, 2020, 25(13): 23-26.
- JIAO H J, DU S X, LIAO C Y, et al. Research on intelligent recommendation technology for official document processing based on TFIDF algorithm[J]. Office Informatization, 2020, 25(13): 23-26.
- [7] 伍舒婷. 基于深度学习的政府公文智能分类技术研究[D]. 成都: 电子科技大学, 2020.
- WU S T. Research on intelligent classification technology for government documents based on deep learning[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
- [8] 黄永, 陆伟, 程齐凯. 学术文本的结构功能识别: 基于章节内容的识别[J]. 情报学报, 2016, 35(3): 293-300.
- HUANG Y, LU W, CHENG Q K. The structure function recognition of academic text—chapter content based recognition[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(3): 293-300.
- [9] 黄永, 陆伟, 程齐凯, 等. 学术文本的结构功能识别: 基于段落的识别[J]. 情报学报, 2016, 35(5): 530-538.
- HUANG Y, LU W, CHENG Q K, et al. The structure function recognition of academic text—paragraph-based recognition[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(5): 530-538.
- [10] GASPARETTO A, MARCUZZO M, ZANGARI A, et al. A survey on text classification algorithms: from text to predictions[J]. Information, 2022, 13(2): 83.
- [11] 马嘉荣, 董珍兰. 浅析智能化工具对公文写作的积极影响: 以WPS智能写作为例[J]. 秘书之友, 2021(10): 28-31.
- MA J R, DONG Z L. Analysis of the positive impact of intelligent tools on official document writing: take WPS intelligent writing as an example[J]. Secretary's Companion, 2021(10): 28-31.
- [12] 柏峰, 李宁, 施运梅. 一种政府公文智能辅助写作系统[J]. 北京信息科技大学学报(自然科学版), 2021(5): 28-32.
- BAI F, LI N, SHI Y M. A government document intelligent assisted writing system[J]. Journal of Beijing University of Information Technology: Natural Science Edition [J]. 2021 (5): 28-32.
- [13] 杨锐, 陈伟, 何涛, 等. 融合主题信息的卷积神经网络文本分类方法研究[J]. 现代情报, 2020, 40(4): 42-49.
- YANG R, CHEN W, HE T, et al. Text classification method based on convolutional neural network using topic information[J]. Journal of Modern Information, 2020, 40(4): 42-49.
- [14] 陈思. 基于深度学习的电子公文智能审查[D]. 武汉: 华中科技大学, 2021.
- CHEN S. Intelligent review of electronic documents based on deep learning[D]. Wuhan: Huazhong University of Science and Technology, 2021.
- [15] 夏义堃. 政府数据治理的国际经验与启示[J]. 信息资源管理学报, 2018, 8(3): 64-72, 101.
- XIA Y K. International experiences and implications on government data governance[J]. Journal of Information Resources Management, 2018, 8(3): 64-72, 101.
- [16] ROCCHIO J. Relevance feedback in information retrieval[J]. Computer Science, 2000: 313-323.
- [17] PARTALAS I, KOSMOPOULOS A, BASKIOTIS N, et al. Lshtc: a benchmark for large-scale text classification[J]. arXiv preprint, 2015, arXiv: 1503.08581.
- [18] ALI T, ASGHAR S. Multi-label scientific document classification[J]. Journal of Internet Technology, 2018, 19: 1707-1716.
- [19] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

作者简介



金加和 (1965-), 男, 浙江省数据开放融合关键技术研究重点实验室副主任, 浙江省大数据发展中心主任、正高级工程师, 浙江省政务服务标准化技术委员会副秘书长, 主要研究方向为数字政府、公共数据、数据治理等。



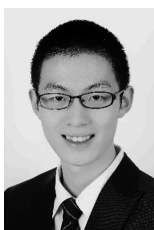
施筱玲 (1983-), 女, 浙江省数据开放融合关键技术研究重点实验室成员, 浙江省大数据发展中心工程师, 主要研究方向为数字政府、数据融合应用、数据标准等。



徐峰 (1992-), 男, 浙江省高层次人才服务中心工程师, 主要研究方向为政务大数据应用、人才大数据分析等。



叶红叶 (1981-), 女, 浙江省数据开放融合关键技术研究重点实验室成员, 浙江省大数据发展中心管理人员, 主要研究方向为数字政府、三农大数据应用等。



俞巍滔 (1989-), 男, 浙江省数据开放融合关键技术研究重点实验室成员, 浙江省大数据发展中心工程师, 主要研究方向为数据管理、数据开发利用、公共数据授权运营等。

收稿日期: 2024-03-15

通信作者: 俞巍滔, yuwt@zj.gov.cn

基金项目: 浙江省2024年度“尖兵领雁”研发攻关计划项目 (No.2024C01212)

Foundation Item: 2024 Key R&D Program of Zhejiang Province (No.2024C01212)