

基于大数据的出租汽车信息资源规划研究

张震, 孔帅华, 彭坤

(郑州大学电气工程学院, 河南 郑州 450001)
✉zhangzhen66@126.com; 1985357098@qq.com; 895250656@qq.com



摘要: 针对多年来出租汽车行业存在的数据冗余、结构混乱等问题, 本文立足行业现状, 以城市出租汽车行业监管服务现实需求为导向, 依托城市政务云及电子政务网络, 利用大数据、物联网、云计算、多媒体和人工智能等技术, 建立出租汽车行业信息库, 实现对出租汽车行业信息的整合和管理, 为未来基于行业信息资源进行的大数据应用分析提供支撑。运行表明, 大数据对出租汽车行业信息资源规划研究具有一定的指导意义, 对提高城市出租汽车的运营效率, 缓解城市交通拥挤具有重要意义。

关键词: 大数据; 出租汽车; 信息资源规划; 数据中心

中图分类号: TP399 **文献标识码:** A

Research on Taxi Information Resource Planning based on Big Data

ZHANG Zhen, KONG Shuaihua, PENG Kun

(College of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China)
✉zhangzhen66@126.com; 1985357098@qq.com; 895250656@qq.com

Abstract: This paper proposes to establish a taxi industry database in order to solve problems of data redundancy and structural confusion in taxi industry over the years. This database is expected to realize the information integration and management of taxi industry, and provide support for future big data application analysis based on industry information resources. Taking advantages of city government cloud and E-government network, the proposed database is designed to meet the needs of urban taxi industry supervision. It uses a variety of technologies, such as big data, Internet of Things, Cloud computing, multimedia, artificial intelligence and so on. Operation shows that big data plays a significant role in planning information resources, improving operating efficiency of urban taxis and alleviating traffic congestion.

Keywords: big data; taxi; information resource planning; data center

1 引言(Introduction)

随着城镇化的不断发展, 城市占地面积和人口快速增加, 城市出租汽车的体量迅速增加, 在城市客运方面的占比日益增加, 出租汽车已经逐渐成为城市客运交通体系的重要组成部分, 并在促进城市经济发展、方便人们出行、扩大就业等方面做出了积极贡献^[1]。然而出租汽车行业也暴露出从业人员素质参差不齐、假牌套牌现象时有发生、传统企业经营模式不足、整体服务水平一般等问题, 严重影响了出租车行业的发展^[2]。因此, 依靠大数据技术和现代化的管理手段, 对

出租汽车行业信息资源规划的研究成为一项急迫的任务, 对于提高出租汽车行业的管理水平, 解决出租汽车行业存在的问题, 促进汽车行业健康发展具有重要意义^[3]。

2 信息资源规划(Information resource planning)

基于出租汽车行业相关数据元标准规范定义的业务对象与属性, 结合出租汽车行业各层面现有系统的应用实践, 以城市出租汽车行业和相关规范为指导思想, 对信息资源进行自上而下的统一数据规划。信息资源规划将行业数据按照业务本质特征进行组织, 将出租汽车行业多年来积累的结构混

乱、数据繁杂、不合理的“数据库”进行规划组织，为出租汽车行业建设提供一个稳定的数据环境^[4]。

信息资源规划的目的是站在出租汽车行业资源规划管理的高度，组织规划出租汽车行业业务和数据模型，以及规划行业业务和数据构架，明确数据内容及其关系，为形成出租汽车行业“一数一源”的基础性和整体性数据资源中心奠定了坚实基础^[5]。

2.1 信息资源规划原则

(1)命名规范

遵循国家或出租汽车行业数据标准的命名规范。

(2)设计规范

按照数据库的规范化标准，遵循第三范式的内容要求进行库、表的设计。

(3)统一数据模型

均采用整体统一的数据模型，并共享所有的公共数据模型。

(4)公共代码

按照出租汽车行业编码规范要求，以保证将来业务数据变革时，行业内部相关系统的数据资源保持一致。

(5)元数据管理

实现对数据采集、交换规则、存储规则等的管理，并对元数据采集信息、交换规则进行修正。采用元数据管理实现对数据结构的全局管控^[6]。

2.2 信息资源规划思路和内容

深化贯彻国家出租汽车优先发展战略，以提升服务水平为宗旨，以提高运营效率和管理水平为重点，以加强出租汽车运行安全保障为关键，抓住大数据、物联网、云计算、人工智能等新技术快速发展的重大机遇，按照“深化数据建设，提高服务水平，建立高效机制”的思想，加强新技术和创新应用，通过先行试点再推广应用的模式，确保满足近期需求，并兼顾长远发展^[7]。

(1)整体规划，建立科学的信息资源规划体系结构

建立科学的信息资源规划体系结构，确保业内人员、开发人员等的管理和维护，以及让使用信息资源的用户从全局上掌握信息资源，便捷、快速地利用数据资源，准确有效地管理、维护数据资源。信息资源规划体系结构包括数据体系指标、数据模型等数据标准化方面的内容，也包括数据采集及管理方面的内容。

(2)数据标准化

数据标准化是数据交换、利用和共享的前提条件。数据标准化包括多方面的内容，既包括数据库结构、数据元、数据指标等数据自身的标准化，也包括元数据的标准和采集数据的标准等工作内容。为了应对将来的工作变化，应建立全局的、多方面的信息资源标准体系。

(3)采用分布采集、集中处理、集中交换的建设模式

分布采集：数据中心根据数据的不同特征，及时把来自不同格式和不同存储方式的元数据采集到数据中心。

集中处理：使用ETL工具软件和开发的处理程序对数据进行清洗、转换的过程，集中在数据交换区中进行。

集中交换：对外交换数据统一由数据中心进行整体规划，进行交换策略、方式的管理，使数据交换过程规范化、

标准化^[8]。

2.3 信息资源建设流程

出租汽车信息资源建设流程如图1所示。

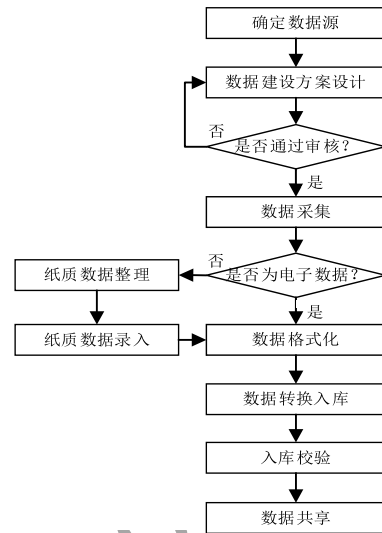


图1 信息资源建设流程

Fig.1 Information resource construction process

(1)确定数据源

对各业务系统和数据资源进行梳理，确定需要采集的数据源。

(2)数据建设方案设计

按照数据源的相关信息，设计一整套数据建设处理方法，包括数据采集、数据整理、数据录入、数据格式化、数据转换入库、入库校验和数据共享等方案^[9]。

(3)数据审核校验

由各相关部门共同审阅数据建设方案，主要对系统采集数据的内容规范性、完整性、有效性和可行性进行校验审核。

(4)数据采集

按照设计的数据建设方案，由专门部门和专业人员在规定时间内对规定数据内容进行采集。

(5)纸质数据整理

当数据资源为纸质资料时，遵循系统数据要求，录入前对纸质数据进行归类，明确录入纸质数据的内容、统计口径、历史口径等相关工作。

(6)纸质数据录入

依据整理好的纸质数据，遵循有关标准，明确录入数据的要求，由专业人员进行数据的录入。

(7)数据格式化

对采集的电子数据或录入的纸质数据，按照系统要求进行数据转换，由专门部门、专门人员，按照要求进行分类、标准化和格式化。

(8)数据转换入库

通过现有的转换程序或编写相应的转换程序，将处理后的数据信息转换入库。

(9)入库校验

对转换入库数据的整体性、精准性和安全性进行校验，保证转换入库的数据规范化、标准化。

(10)数据共享

在规定时间内，按照跨部门、跨行业的管理和服务需求，对所需的数据进行共享。

2.4 出租汽车元数据标准规范

元数据是关于出租汽车信息资源标识、内容、质量、状况及其他有关特征的全面描述，是出租汽车行业数据标准的主要体现形式，其目的是为帮助发现数据、管理数据、使用数据和共享数据提供服务，元数据已成为当今世界范围内信息资源组织和共享必不可少的重要方法和工具。

基础信息资源数据结构类型复杂，对其进行描述的数据本身也是复杂的，因此相对于基础信息资源库还需要建立一个独立的元数据库，并实现好元数据与其表达的实际数据集或数据集系列之间的链接。元数据库直接对外开放，提供多种途径供第三方系统或最终用户查询，使用者通过查询和检索这些组织良好、结构标准的元数据，快速定位到自己需要的信息，进而在经过验证后被导向到实际的数据集服务处，最终检索到自己感兴趣的信息。因此，一个定义、维护良好的元数据库标准是进行出租汽车基础信息资源有效共享的基础，也是实现有效多维数据分析和数据挖掘的前提。

2.5 出租汽车数据标准指标体系

数据标准应从全面的信息资源范围进行定义，涉及的数据资源包含多个“数据集系列”和“数据集”。根据目前出租汽车行业信息化的现状，资源整合库的设计是该指标体系的一个子集合。随着信息化建设工作的不断完善，在今后可以实现指标体系中的所有数据集。

3 信息资源数据中心设计(Design of information resource data center)

数据中心设计的主要目的是服务于城市出租汽车行业，实现行业信息资源的整合，为数据信息资源规划奠定基础，建立出租汽车行业基础信息数据库，实现城市出租汽车行业GPS/北斗数据和营运数据的整合，以及基础数据的管理，建成出租汽车行业数据中心总体框架^[10]。

3.1 数据中心总体架构

数据中心是城市出租汽车行业信息管理系统的信息资源中心，是系统的重要组成部分。其重要功能包括：数据汇集、数据交换、车载设备、呼叫中心实时数据交换、数据共享、存储和分析等。数据信息资源总体架构如图2所示。

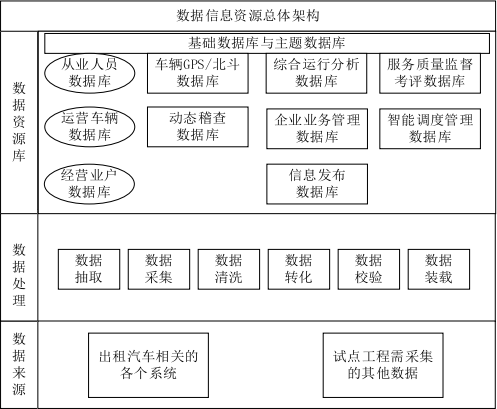


图2 数据信息资源总体架构

Fig.2 Overall structure of data information resources

3.2 数据中心功能

数据中心的重要功能包括数据采集、存储和交换。数据整合与服务功能图如图3所示。

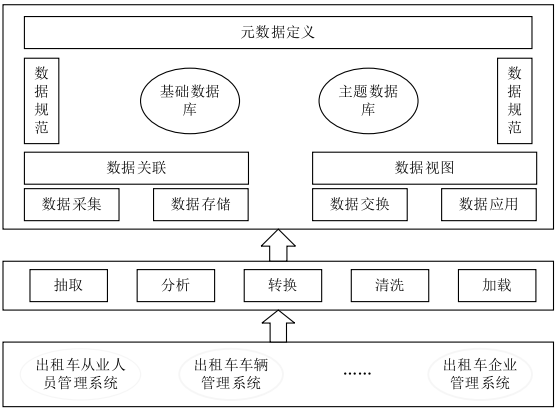


图3 数据整合与服务功能

Fig.3 Data integration and service functions

数据采集：完成多元化的、可应对的数据抽取、分析、转换、清洗和加载的应用平台，将来自不同存储方式和不同格式的数据进行组织分类后录入数据库。

数据存储：结合系统需求中数据保存周期的要求，对应用后的数据生命周期管理策略予以明确，如数据在线保存周期、历史数据清理和备份方式、历史业务数据存档策略等。

数据交换：实现各个业务系统之间的数据共享和交换，对各应用数据资源进行整合和共享访问，防止各应用系统产生数据孤立的局面，完成数据的共享和数据的交换，协同各业务系统之间的数据运行。

3.3 基础数据库设计

根据出租汽车行业信息基础支撑的需要，本文研究规划的基础数据库分别为驾驶员数据库、车辆数据库和企业数据库，其中主要库表如表1和表2所示。

(1)出租汽车驾驶员数据库

表1 出租汽车驾驶员数据库内容列表

Tab.1 Taxi driver database content list

序号	数据类别	数据项	更新频率
1	驾驶员基本信息	姓名、性别、出生日期、身份证号码、照片、民族、联系电话、联系地址、户籍地址、文化程度、从业资格类别、从业资格证号、证照状态、从业资格证初领时间、从业资格证发证日期、证件有效期、发证机关、驾驶证号、准驾车型、初领驾驶证日期、所属企业组织机构代码	实时更新
2	驾驶员审核信息	审核机构代码、证件类别、审核结果	一年更新一次
3	驾驶员变更信息	变更前机构代码、变更后机构代码、变更内容	实时更新

(2)出租汽车车辆数据库

表2 出租汽车车辆数据库内容列表

Tab.2 Content list of taxi vehicle database

序号	数据类别	数据项	更新频率
1	车辆基本信息	车辆牌号、车辆品牌、车牌颜色、车身颜色、车辆类型、厂牌型号、发动机号、出厂日期等	实时更新
2	车辆营运基本信息	道路运输证号、道路运输证发证机构、经营权类型、经营范围、车辆营运状态等	一年更新一次
3	车辆审核信息	审核日期、审核记录、审核结果等	一年更新一次
4	车辆转籍信息	转籍类型、转籍记录、转籍日期、证明材料等	实时更新
5	车辆报废信息	报废类型、报废原因等	实时更新
6	车辆稽查信息	违法日期、违法地点、违法类型、处罚决定日期、处罚金额、执法机构等	实时更新

3.4 主题数据库设计

主题数据库包括综合运行分析数据库、服务质量监督考评数据库、动态稽查数据库、车辆GPS/北斗数据库和信息发布数据库^[11]，其中主要的库表如表3至表5所示。

(1)综合运行分析数据库

表3 综合运行分析数据库内容列表

Tab.3 Content list of comprehensive operation analysis database

序号	数据类别	数据项	更新频率
1	车辆营运信息	车辆编码、企业编码、从业资格证件号、上下车地点经纬度、等候时间、上下车时间、车辆行驶速度信息、车辆状态信息、附加费、交易记录等	实时更新
2	车辆行驶信息	车辆编码、行驶速度、营运里程、空驶里程、车辆位置经纬度、车辆状态信息等	实时更新

(2)服务质量监督考评数据库

表4 服务质量监督考评数据库内容列表

Tab.4 Content list of service quality supervision and evaluation database

序号	数据类别	数据项	更新频率
1	投诉信息	车辆编码、企业编码、从业资格证件号、车辆速度信息、车辆位置经纬度、上下车时间、交易记录、投诉时间、投诉内容、投诉备注等	实时更新
2	单位资质信息	单位资质编号、单位资质名称	实时更新
3	信用标准信息	信息标准编号、信用指标编号、信用标准权重、信用标准描述	实时更新
4	信用指标类别信息	信用指标编号、信用指标名称	实时更新

(3)动态稽查信息数据库

表5 动态稽查信息数据库内容列表

Tab.5 Content list of dynamic audit information database

序号	数据类别	数据项	更新频率
1	车辆行驶信息	车辆编码、行驶速度、营运里程、空驶里程、车辆状态信息、位置经纬度等	实时更新
2	车辆黑名单信息	驾驶人员编码、车辆编码、车辆牌号、车牌颜色、车身颜色、车辆类型、车辆品牌等	一周更新一次

4 结论(Conclusion)

科学的出租汽车信息资源规划可以缓解城市交通拥挤，满足了广大人民群众出行的需求，助推城市出租汽车行业形成规范有序、健康文明、市场净化、协调发展的新格局，提高了出租汽车行业监管及服务能力。本文从出租汽车信息资源规划思路、内容、建设流程等方面进行了分析研究，并且对信息资源数据中心设计进行了阐述，给出了数据信息资源总体架构，为出租汽车行业管理提供了依据。

参考文献(References)

[1] 国务院关于印发促进大数据发展行动纲要的通知.国发〔2015〕50号.

[2] 刘响.城市出租汽车市场运营博弈策略及其影响研究[D].哈尔滨:哈尔滨工业大学,2018.

[3] 杨英俊.城市出租汽车服务管理关键技术研究[D].西安:长安大学,2013.

[4] A G W, B E T. Using "Big Data" to understand the impacts of Uber on taxis in New York City[J]. Travel Behaviour and Society, 2020(22):94-107.

[5] 刘增.城市出租汽车监管与服务平台的设计实现[D].哈尔滨:黑龙江大学,2015.

[6] 蔡筹皓.基于首都机场出租车的信息采集与处理研究[J].产业创新研究,2020(16):111-113.

[7] 刘荣.出租车合理规模研究与应用[D].长沙:长沙理工大学,2013.

[8] 焦红红.城市出租汽车决策支持方法研究及系统开发[D].西安:长安大学,2014.

[9] 刘文婧,韩建业.构建电子数据归档管理系统的实践[J].中国国情国力,2018(08):76-78.

[10] Xu W, Huang Y. Mining Urban Congestion Evolution Characteristics Based on Taxi GPS Trajectories[J]. American Journal of Traffic and Transportation Engineering, 2020, 5(1):1.

[11] 陈睿,刘伟,孟庆久.信息资源规划中主题数据库规划研究与应用[J].情报杂志,2006,25(2):136-137.

作者简介:

张震(1966-),男,博士,教授.研究领域:信息与通信工程,大数据技术.

孔帅华(1995-),男,硕士生.研究领域:计算机应用,大数据技术.

彭坤(1996-),男,硕士生.研究领域:计算机应用,大数据技术.