

我国电动汽车数据库建设的几个问题

张旭明¹ 赵立金² 葛莹莹¹

(1. 北京中汽虹图汽车技术开发有限公司, 北京 100038; 2. 中国汽车工程学会, 北京 100055)

摘要: 从数据库设计、数据采集及数据库维护及更新等3个方面讨论在建立我国电动汽车专业数据库过程中所遇到的问题及提出的解决方案, 这对于电动汽车专用数据库建设具有一定的借鉴意义。

关键词: 电动汽车; 数据库; 车型; 动力电池; 驱动电机

中图分类号: T19

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2015.01.016

Some Problems for Building of Electric Vehicle Special Database on Our Country

Zhang Xuming¹, Zhao Lijin², Ge Yingying¹

(1. Beijing Bright Future Auto Industry Consulting Company, 100038; 2. Society of Automotive Engineers of China, 100055)

Abstract: Problems of Electric Vehicle (EV) Special Database building and solving scheme are discussed by structure design, data collection and process, and database maintenance, there out valuable references are provided to EV special database build In this paper, on our country.

Keywords: electric vehicle(EV), database, vehicle model, power battery, traction motor

我国从“九五”开始就启动了电动汽车的研发工作, 各研究机构和单位虽已积累了大量数据, 但未能整合形成统一有效的专用数据库, 尤其是车型、关键零部件、示范推广等类型的专用数据更是缺少。目前, 行业内的部分电动汽车网站可以提供一部分数据查询功能, 但主要以资讯信息为主, 总体上尚未形成有效全面的电动汽车决策数据库系统。另外, 目前的汽车数据库主要针对传统汽车(主要是内燃机汽车, 如汽油车、柴油车), 没有针对电动汽车的技术特点及关键

零部件的发展特点进行优化。例如, 没有针对电动汽车纯电行驶、动力电池、驱动电机的字段。而且, 在字段设计时, 由于是资讯信息为主, 即以文本信息为主, 这种设计是针对查阅, 对于后续数据分析及处理不是很方便, 如针对电动汽车续航里程、动力电池重要的能量密度指标等, 无法进行直观对比。因此, 建立电动汽车领域专用数据库系统具有重要意义。

本文将从数据库设计、数据采集、数据库的维护及更新等3方面讨论我国在建设电动汽车专

作者简介: 张旭明(1972-), 男, 北京中汽虹图汽车技术开发有限公司, 研究方向: 汽车技术、政策, 新能源汽车、替代能源及智能交通; 赵立金*(1985-), 中国汽车工程学会, 研究方向: 汽车产业战略规划、技术研究及重大技术专项管理等; 葛莹莹(1988-), 女, 北京中汽虹图汽车技术开发有限公司, 主要研究方向: 汽车技术及相关政策。

基金项目: 国家科技支撑计划课题“电动汽车专题数据库建设”(2013BAG06B02); 国家科技支撑计划课题“电动汽车技术预测及决策支持应用示范”(2013BAG06B04)。

收稿日期: 2014年5月19日。

用数据库过程中所遇到的问题及解决方案。

1 数据库类型与数据库结构

考虑到使用的需求、建设数据库的主要目的和数据库功能以及电动汽车除了关键零部件（包括电池和驱动电机）之外其他零部件因与传统内燃机汽车比较接近等因素。笔者认为，应建设车型数据库、关键零部件数据库（包括电池数据库、电机数据库）和产业化数据库。其中，车型数据库主要跟踪主要车辆产品的技术情况，将包括纯电动汽车车型及插电式混合动力车型；电池数据库收集主要电池产品技术参数；电机数据库收集主要电机产品的技术参数；产业数据库则针对零部件厂商的产能规划及合作、配置关系。

在数据库设计中最关键的是数据库的结构。为了提供可用性强的数据支撑，要根据实际需求来确定数据库的结构，以“动力电池产业化数据库”为例，数据库构成部分如表1所示。这种结构有助于形成各项目的对比，突出重要部分数据，根据各厂商的不同产品、产能发展和规划以及不同年代的产值进行评价，并对比各厂商之间的相互配套关系，了解上下游企业之间的合作关系，以此来评估产业链的完整性以及供应商和整车厂的成熟程度。

以前的评价通常是具有主观性、定性的评价。通过建立完整的数据库，希望达到定量、客观的评测，因此在数据库的结构和采集的项目上要求更加严格。原有的数据库结构也存在不完善的地方，例如在“车型数据库”中，主要性能参数一栏列举了“快充时间”“充电时间（慢充，

220V）”这两项（表2），这种结构有缺陷，需要改进。笔者建议在后边加一栏“充电模式”，这样可以更直观地对比不同车型的充电模式，进行总结分类。建议增加数值型项目，使得数据更加直观，便于对比各项目间的参数，有利于做出客观的评价。建议建立统一的受控词以达到数据库的结构完整，内容规范，便于整理和检索。可以参照已形成的标准，通过原有的模式制定电动汽车的受控词体系。在源文件中对产品类别和主要产品的表述并没有规范成统一的受控词，如在电池产业化数据库中，可分为两个等级，产品类别中为一级标题，主要产品为二级标题，在录入整理的过程中能规范到二级标题的尽量使用二级标题，否则使用一级标题来表示，如表3所示。由于不可能列举所有受控词，同时避免过于细分影响技术类别的判定，因此电池数据库及电机数据库对于等级只划分为二级，而且只是针对目前发展较快或者关注较多的技术。对于应用较少的技术，如镍氢电池主要应用于混合动力汽车，铅酸电池很少用作动力电池，燃料电池在车上的应用主要为质子交换膜燃料电池，因此对于这些都没有进行细分。随着技术进步，如果有新的电池技术出现，可以进行扩充。这在数据库的设计中应该加以考虑。

通过改进，可分清概念的等级结构，有助于数据库的规范化，以方便检索。在数据库设计过程中要考虑数据库的可扩展性，例如在结构设计中要为可能的技术进步提供可扩展的空间。在设计数据库字段时，要考虑便于后续处理。电动汽车数据库的功能就是为决策提供科学、客观的支

表1 动力电池产业化数据库

厂商	国别	产品类别	主要产品	现有产能	产能规划	战略合作	配套关系	注册资金	股份构成	公司所在城市
----	----	------	------	------	------	------	------	------	------	--------

表2 电动汽车车型数据库性能参数

主要性能参数											
纯电 续驶 里程	最大 续驶 里程	最高 车速	加速性能 (0— 100km/h)	加速性能 (0—60km/h)	燃料经济 性(电) 工况法	燃料经济性 (电)等速, 60km/h	燃油经济性 (油、气等)	充电时间 (慢充,220V)	快充 时间	整车 质量	车身 尺寸

表3 动力电池及驱动电机分类

动力电池		动力电机	
一级	二级	一级	二级
锂电池	磷酸铁锂	交流电机	永磁同步
	锰酸锂		磁阻电机
	钛酸锂		交流感应（异步）
	三元材料		交流同步
	磷酸铁钒锂		
	其他		
镍氢电池		直流电机	直流电机
铅酸电池			
超级电容			
燃料电池			

撑，因此需要更多量化、系统化的分析处理。在设计字段时，尽量采用数值型数据，以便进行平均、最大值等数学计算。

2 数据采集及加工

数据采集及加工通常包括确定数据来源、数据采集、数据整理及清洗、数据建库及检查等流程。

在数据的采集方面，汽车产业的数据发布没有统一的格式及渠道，因此数据库主要是以人工采集为主，通过多种渠道寻找可用数据，主要包括企业官方网站、商业网站和政府渠道，如，车型数据库和关键零部件数据库中的数据多来源于各厂商官网，包括像“比亚迪官网”“福田汽车官网”“第一汽车制造厂”等，大部分数据来源于“全球汽车产业平台”；示范运营数据库中的数据来源于“全球汽车产业平台”以及各科委的网站等。因此，由于数据的来源不同，导致数据格式各异，较难统一。数据库整理过程繁琐的根源是信息渠道不同，直接导致数据不规范。部分数据直接来源于各厂商网站，也有部分信息来源于国家各部委网站。两种渠道数据的可靠性无从考证，在两类数据出现不一致时，无法判别哪一类数据是正确的，因此数据录入时便产生了一定的误差，并且各类数据源的规范标准不相同，所用格式各式各样，较难统一成一种格式。电动汽车专用数据库所录入的数据，源自于各网站统计的

结果，不能保证信息的全面性，存在部分数据缺失的情况，而且各网站的数据更新频率不同，如果这类数据是比较重要，则需要通过专家协商来补齐数据。此外，还可以借鉴其他行业的数据采集方式加强电动汽车数据的采集，例如进行网页抓取的方式，这在其他课题中有所涉及。

在数据的加工方面，由于没有统一的规范标准，因此采集的数据在加工处理时会遇到各种问题。例如，采集到的数据有的为区间值，有的则是一表多值或者采用文字描述。这导致了数据库的不可用性，因此要将所有的区间值都采用平均数的算法改成单一值，一表多值的数据分成多个字段表示，涉及产能、性能参数的数据，既有数值型，也有文本型，这种文本冗长多余，在数据库中是无法使用的。因此要筛选排除无效字段，如在关键零部件数据库中，主要产品型号有多个值，现有产能和产能规划为文字描述的形式，为保证数据可用，要将主要产品中的多个型号统一改成每个单元格表示一个型号的格式，并分成多个表格来表示，并将产能规划中的文本改成数值形式，删除不可用的文本，以保证数据库的清晰简洁。

对于数据库中所涉及的机构要统一规范。产业化数据库中的配套关系是为了准确评估产业链的完整性以及评估供应商、整车厂的成熟程度。这一部分涉及的企业绝大部分用的是简称，此部分需要提供全称格式。但是有的单位所用字段十分简洁，单凭搜索其简称并不能找到对应的全称，不能确定哪一个全称是源文件中所要列举的对象，如“东风汽车”是一个公司简称，视不同情况，可能指代东风汽车公司（总公司）、东风汽车集团股份有限公司（香港上市公司）、东风汽车有限公司（中日合资）、东风汽车股份有限公司（上海上市公司）等公司中的某一家。如果这些数据不能规范成统一单位名称，必将导致数据的不可用性，因此需将原文中所有的单位统一规范，进行机构归一处理，避免出现一个单位名称两种表示方式，从而避免将来进行统计分析时出现漏项的可能，还可以在产业化数据库中单位全称的

后面添加一栏“简称”，这样既能方便查看，又避免了单纯用简称带来的误解。另外，由于在多个数据库中都存在机构，因此要建立统一的机构数据库进行规范管理。这就要求各数据库中的机构进行统一规范。

尽量统一规范计算单位，关键零部件数据库和产业化数据库中出现了多种计算单位，格式各异，如产能单位，表示方法多样，有“KW/年、kw/y、kw/日、kw/月”等，产能规划中既有“元/年”，也有“美元/年”，有些单位可以统一规范，有些则无法统一，因此考虑数据的可用性及可对比性并进行处理，需要将产能单位统一改为KW/年，其他数据根据重要性可删除或者换算成统一单位。同时，可以增加原始数据字段，对原始数据进行保留，以便查询。

在数据加工处理的过程中以及后续的应用中，现有的电动汽车数据还很不完善，可以利用的商业数据库还不够多，这与电动汽车尚处于起步阶段有关。

3 数据的更新与维护

在数据库整理过程中，必不可少的工作是数据的更新与维护，例如车型数据库中，新车型不断涌现，需要不断更新添加车型。一方面可以根据新车型集中发布的事件（如有影响力的大型车展）进行定期更新，或者定期对主要厂商的官方网站进行扫描；另一方面可以把整理好的信息进行二次处理，使得数据的准确性高、可用性强，并定期进行数据的更新，补充新数据，替换旧数据，保证数据库随着科技迅速发展做到最新。只有这样，电动汽车产业规划的评价体系才能更加规范，评价的结果才能更加准确，误差更小，有利于对电动汽车的发展做出更精准的预测。

4 结论

电动汽车数据库开发将直接用于评价电动汽车的发展，为技术预测与决策提供有效的、科学的、客观的数据支撑，因此为了方便日后的检索

和应用，需要把各类数据统一整理规范。但是，在加工处理这些数据的过程中产生了很多问题，如何解决这些问题。制订统一的标准可以更好地建设、维护及应用数据库。对于专业数据库，既要考虑专业性，即电动汽车行业人员对数据库的使用需求，也要考虑数据库的通用性，即通用数据库的特点以及建设规范。

通过建立完善的数据结构，掌握良好的数据设计和采集方法，解决加工处理过程中发现的问题，有效地管理数据库，从而改变数据通用性差、不便于移植、不便于更新的缺陷，并制订统一的标准，为日后汽车产业化数据库的规范化打下基础，提高数据的准确性，方便与各种应用程序或应用系统接口，以更加有效地使用数据库中的数据。

在电动汽车这种专业数据库的建设中，我们与中国科学技术信息研究所开展了紧密的合作。汽车专业与信息技术行业的跨行业合作，既可保证数据库满足专业的应用需求，也可保证数据库本身规范化。

参考文献

- [1] 中国汽车技术研究中心. 2013 节能与新能源汽车年鉴 [M]. 北京：中国经济出版社，2013.
- [2] 国务院发展研究中心产业经济研究部，等. 中国汽车产业发展报告 [M]. 北京：社会科学文献出版社，2012.
- [3] 中国汽车工程学会，等. 中国战略性新兴产业研究与发展 [M]. 北京：机械工业出版社，2013.
- [4] 欧阳明高. 汽车新型动力：系统、模型与控制 [M]. 北京：清华大学出版社，2008.
- [5] Alvaro Masias, Kent Snyder, Ted Miller. Automaker Energy Storage Needs for Electric Vehicles[C]// Proceedings of the FISITA 2012 WorldAutomotive Congress, 2012.
- [6] 郭建龙, 陈世元. 电动汽车驱动用电机的选择 [J]. 汽车电器, 2007(1): 9-12.
- [7] 弗罗斯特. 数据库设计与开发 [M]. 北京：清华大学出版社，2007.