

聚变实验装置数据库设计^①



蔡 维^{1,2}, 尚雷明¹, 杨子辉¹, 石志勇³, 郝丽娟¹, 胡丽琴¹

¹(中国科学院 核能安全技术研究所 中子输运理论与辐射安全重点实验室, 合肥 230031)

²(中国科学技术大学, 合肥 230027)

³(中国人民解放军 96901 部队, 北京 100094)

通讯作者: 尚雷明, E-mail: leiming.shang@fds.org.cn

摘 要: 聚变实验装置是开展聚变堆设计及实验研究的重要平台. 但目前聚变领域缺乏集成的聚变实验装置数据库及有效的装置性能分析工具, 限制了科研人员聚变设计研究效率的提升. 本文设计并研发了一套聚变实验装置数据库, 解决了多源数据规整与异构数据存储等问题, 实现了典型聚变实验装置的相关参数与资料的统一入库管理, 以及多字段智能检索与多维可视化比对分析, 为聚变研究人员提供可靠, 便利的聚变实验装置数据服务.

关键词: 核聚变; 聚变实验装置; 数据库

引用格式: 蔡维, 尚雷明, 杨子辉, 石志勇, 郝丽娟, 胡丽琴. 聚变实验装置数据库设计. 计算机系统应用, 2020, 29(10): 109-113. <http://www.c-s-a.org.cn/1003-3254/7661.html>

Design of Fusion Experimental Device Database

CAI Wei^{1,2}, SHANG Lei-Ming¹, YANG Zi-Hui¹, SHI Zhi-Yong³, HAO Li-Juan¹, HU Li-Qin¹

¹(Key Laboratory of Neutronics and Radiation Safety, Institute of Nuclear Energy Safety Technology, Chinese Academy of Sciences, Hefei 230031, China)

²(University of Science and Technology of China, Hefei 230031, China)

³(Troops 96901 of PLA, Beijing 100094, China)

Abstract: Fusion experimental device is an important platform for fusion research. The performance parameters of the fusion device have important reference value for the fusion researchers to design the fusion reactor and select the experimental device. However, there is no integrated database for fusion experimental devices, which seriously affects the fusion research efficiency. In this study, a special database of fusion experimental devices was designed and implemented, which solves the problems of multi-source data regularization and heterogeneous data storage, and realizes the unified storage of relevant parameters. The functions of multi-field intelligent retrieval and multi-dimensional comparison analysis are developed, providing reliable and convenient data service for fusion researchers.

Key words: fusion; fusion experiment device; database

聚变实验装置是开展聚变堆设计及实验研究的重要平台^[1], 现有聚变实验装置的参数指标对聚变堆系统的设计创新工作具有重要参考价值^[2]. 但参数指标等数据资料来源多样, 且国内缺少便捷的聚变实验装置数

据检索渠道与可视化对比工具^[3], 国际上也尚未建成针对聚变实验装置的数据库系统, 限制了科研人员的信息获取效率^[4,5]. 因此, 建立一套针对聚变实验装置的专业数据库, 降低聚变实验装置信息获取与分析的成本,

① 基金项目: 中国科学院“十三五”信息化专项科研信息化应用工程项目 (XXH13506-104); 国家基础科学数据共享服务平台 (DKA2017-12-02-XX)

Foundation item: Special Project of Informatization of Chinese Academy of Sciences of Chinese Academy of Sciences during Thirteenth Five-Year Plan (XXH13506-104); National Basic Scientific Data Sharing Service Platform (DKA2017-12-02-XX)

收稿时间: 2020-03-26; 修改时间: 2020-04-24; 采用时间: 2020-05-10; csa 在线出版时间: 2020-09-30

对提高聚变堆设计研究效率具有重要意义。

本文在核安全科技领域云与聚变物理专题数据库的框架下^[6],设计开发了一套聚变实验装置数据库系统,解决了多源异构数据的采集,规整,存储,查询问题,实现了针对聚变实验装置数据的多字段智能检索与多维可视化比对分析功能;收录了国内外主要在运,在建,以及退役的聚变实验装置的基本信息,设计参数,运行参数等数据.为聚变科研人员提供稳定,可靠的装置数据源与方便,快捷的数据服务.

1 数据库系统设计与实现

1.1 系统架构

聚变实验装置数据库的主要建设目标是为用户提供便捷,可靠的数据服务,同时服务的稳定性,可维护性是数据库服务的基本需求,因此数据库系统设计阶段需要考虑框架结构的稳定性,可靠性,同时易于拓展,便于多字段智能检索与多维可视化比对分析工具开发.

为满足上述条件,本文采用成熟稳定的MVC架构^[7]作为基础架构,以保障服务的稳定可靠.同时采用了Vue.js^[8]作为前端框架,以保障功能拓展的便利性.

此外,本文研究设计并构建了针对多源数据的数据规整方法;实现了对参数指标,文档图片等异构数据的统一存储;开发了面向聚变实验装置的多维可视化比对分析工具,为用户提供稳定便捷的综合数据服务.

聚变实验装置数据库的总体结构架构如图1所示.数据库的总体架构分为4层,分别为数据层、数据持久层、业务层、表示层^[9].其中,数据层为底层数据库,进行数据资源的存取管理;数据持久层进行数据库链接管理,保障数据库运行的稳定,可靠;业务层为后台服务程序,处理用户请求;表示层为业务前端,进行用户输入处理,数据可视化展示.

1.2 数据库结构设计

聚变实验装置数据的主要数据类型包括数值数据,文本数据,图片数据,文档数据.对这些异构数据的存储是数据库结构设计的主要问题.

这4类数据中,数值数据和文本数据直接存储于数据库中,图片数据与文档数据则存储于文件服务器中,并将引用链接以数据的形式存储于数据库中.数据库的主要表结构如图2所示.

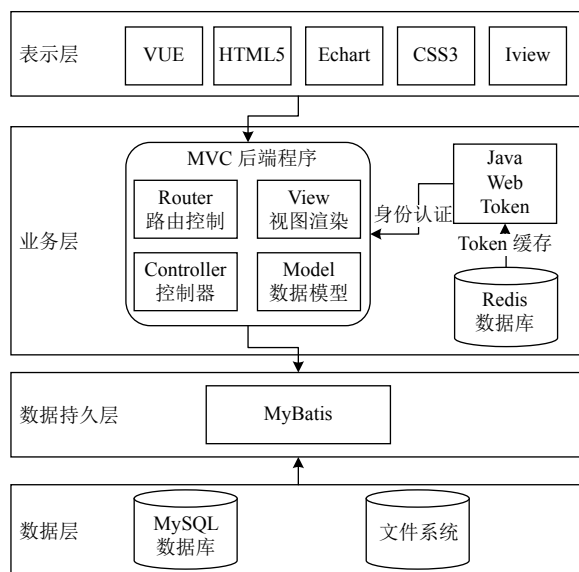


图1 聚变实验装置数据库架构

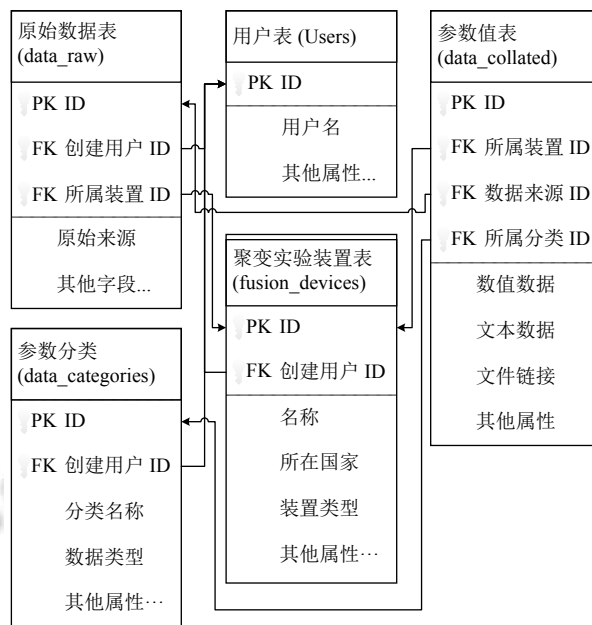


图2 聚变实验装置数据库UML图

聚变实验装置表 (fusion_devices) 存储了聚变实验装置的基本信息,包括名称,所在国家与类型,所在机构等;同时存储了等离子体大半径等常用关键参数,用于提供快速检索服务.

由于聚变实验装置参数具有类型多样,不断更新的特性,针对同一装置往往有不同版本,不同来源,不同类型的参数.如JET装置就从圆截面偏滤器型托卡马克升级为D偏滤器型托卡马克^[10],装置类型,特征均有变化,单一数据结构无法有效存储.

为解决这一问题,聚变实验装置数据库中通过建立单独的参数分类表 (data_categories) 来进行参数类型的管理.装置因升级而产生新类型的参数数据后,管理员可直接通过参数分类表来直接新增新的参数类别,以支持聚变实验装置升级改造产生的参数变动.同时参数分类表作为标准参数名对照表进行数据规整.

参数值表 (data_collated) 中存储经过数据规整后的装置数据,主要字段为数据类型,数值数据值,文本数据值,同时记录管理员审核的相关信息,其中文本数据中包括了图片,文档等参考文献的链接.

1.3 系统实现

聚变实验装置数据库采用前后端分离的 MVC 架构,使用 MySQL 作为基础数据库,Mybatis 作为数据中间件,Vue.js 作为前端框架的技术栈^[11].

数据层采用 MySQL 数据库作为底层数据库管理系统 (RDBMS) 对数据进行管理.同时针对图片,文档等非结构化数据,直接使用文件系统进行存储,管理.

数据持久层采用 MyBatis 作为 DAO (Data Access Object) 组件,以完成对数据库访问的持久化设置^[12],保障数据服务的稳定性,可靠性.

业务层采用 MVC 架构,进行后台服务,处理用户请求及进行模型,视图处理,生成用户视图返回表示层的前端应用.同时,聚变实验装置数据库管理系统中,使用 Redis 数据库^[13] 作为 Token 缓存数据库,处理用户在线状态等信息^[14],实现用户数据上传及管理员身份认证等功能.

表示层前端应用采用 Vue.js 作为基础框架,以此为基础开发实现了多字段智能检索功能的前端控制部分,并对后台检索到的数据进行多维可视化分析.图3为聚变实验装置数据库的首页.



图3 聚变实验装置库首页

2 主要功能与特点

聚变实验装置数据库面向用户提供多字段智能检索,多维数据比对分析功能,能够对聚变实验装置数据进行快速检索并进行可视化的数据分析.

2.1 多字段智能检索

聚变实验装置的数据字段繁多且格式多样,单一条件检索难以定位用户所需的装置数据.为解决这一问题,聚变实验装置数据库向用户提供多字段,多关键词的智能检索功能.

在多字段智能检索页面中,用户可自行添加检索条件,并设置检索条件间约束关系.检索条件中,支持用户指定装置名称,装置所在国家,装置所在单位,装置建成年份等文本关键词检索.同时,针对装置大小半径,中心磁场强度,等离子体最大电流,辅助加热总功率等数值参数,支持用户进行大于、小于、介于等数值条件检索.图4展示了多字段检索功能界面及检索示例.



图4 多字段检索功能

为避免繁琐的检索条件输入和复杂的检索界面,本系统采用通用检索和自定义条件数目来提升检索效率,使用户可根据需求进行单条件简单检索和多维复合检索.此外,聚变实验装置数据库还提供数值,文本的跨类型检索.功能实现上采用了前后端分离,前端条件处理,后端语句查询的方式.具体过程如图5所示.

前端程序中,用户输入,新增的检索条件由 Vue 前端程序进行控制,包括新增条件,数值条件与文本条件的区分,逻辑关系梳理.用户完成条件设置后,整理为包含了所有文本条件及数值条件的条件数组,并形成用于文本检索的 SQL 语句和用户数值检索的检索条件.前端处理完成后,通过 JORN 数据包的形式传送到业务层后台应用.

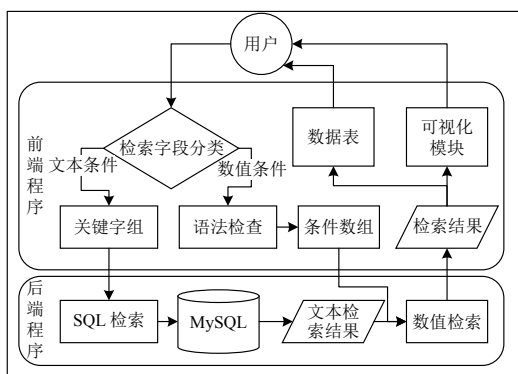


图5 多字段检索功能流程图

后台应用中,首先针对文本条件对 MySQL 数据库进行检索,检索完成后针对返回的结果数组按照数值检索条件进行数值匹配,得到符合用户条件的检索结果,生成视图并返回表示层前端,供用户浏览或进行比对分析。

2.2 多维度可视化比对分析

聚变实验装置的结构多样,参数繁多,且不同装置侧重点不同,为了直观对比不同装置各方面的差异,特征,本系统开发了多维度对比分析功能,支持等离子体大半径,等离子体小半径,最大磁场强度,最大等离子体电流,辅助加热总功率 5 个维度的参数可视化对比功能,向用户提供直观的数据对比工具。

该功能基于 echarts 数据可视化插件进行开发,并针对聚变实验装置的特性,制定了不同的可视化方案,为用户提供了雷达图、平行坐标图、柱状图 3 种数据可视化方案。

通过雷达图,用户可直观了解不同聚变实验装置各个方面的综合差异;平行坐标图则展示了针对某个参数的装置特征;柱状图则是基础视图,供用户选择使用。图 6 为多维可视化对比分析功能的具体示例。

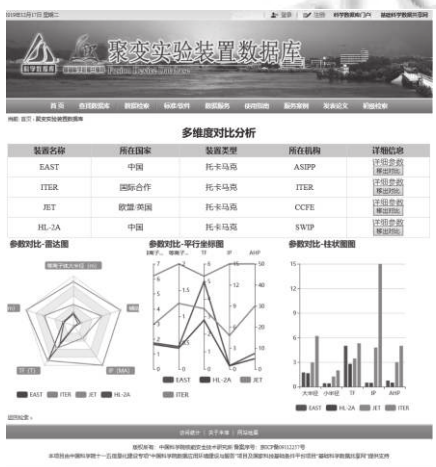


图6 多维可视化对比分析功能示例

2.3 数据上传审核及编程接口

聚变实验装置数据库的初始数据来源于人工收集,整理,校验.为了保证数据的广泛性和及时性,聚变实验装置数据库向用户开放了数据上传功能。

普通用户可通过聚变实验装置数据库系统向数据库中提交新数据,而数据库管理员则可通过管理员账号对数据库中所有数据进行管理,并对聚变实验装置数据库中其他用户上传的数据进行审核操作。

此外,聚变实验装置数据库同时也向用户提供了应用程序编程接口(API),用户可通过 API 进行检索,获取数据库数据等操作,开发第三方应用.关键接口如表 1 所示,其中多字段复合检索功能的关键词 SQLstr 为多字段复合检索中的文本检索条件,而 condition 数组则为数值检索条件组.聚变实验装置数据库中,所有 API 的返回值均为 JSON 数据包,用户可根据相关文档进行处理和使用。

表 1 聚变实验装置库系统 API(部分)

API	功能	参数
simple_search	单关键字简单检索	keyword
complex_search	多字段复合检索	SQLstr, condition[]
detail_doc	拉取详情页面,所有数据以文本形式给出,包括图片,文档链接等	id
detail_para	获取详细参数,以数值方式给出	id

3 数据采集与规整

聚变实验装置数据是典型的多源异构数据.数据采集,数据规整是保障数据库准确性的重要工作内容。

3.1 数据采集

聚变实验装置数据主要来源于各装置所在机构发布的官方文档,论文,报告等文献资料以及网络百科,国际公开数据库等网络渠道.数据结构方面包括了数值数据,文本数据,图片资料等类型。

数据采集过程主要通过人工收集,整理进行,并在数据库中以长文本的方式记录数据名称,数据值,数据标题,数据更新时间等关键数据。

对于网络数据源,如网络百科,机构网站等,在数据收集过程中直接保存网页中所有文本,链接.对于这部分资料,以长文本的形式存储于数据库中,而对其中的图片,视频等数据,则下载到文件服务器中保存,并在数据库中记录存储路径与引用链接。

对于文献数据源,由人工录入相关参数,手动输入

文献中涉及到的装置参数名与参数值,并录入相关的介绍文本等信息.以文本方式存储于数据库中,并保存文献发表时间,内容标题,文献链接等资料.

此外,部分新闻稿件中也会涉及到相关装置的参数值,参数名,对于这些数据源,则作为参考资料存储于原始数据表中,供后续应用开发使用.

3.2 数据规整

原始数据录入后,是以长文本的形式存储于 MySQL 数据表中,而管理员审核的是装置的数值化数据.将文本数值转化为供管理员审核的数值数据这一过程为数据规整的预处理过程.

预处理具体步骤如下:

0) 初始化: 原始文本 RC, 标准参数名 SN, 处理后的参数值 PP.

1) 从 data_raw 中读取未经处理的原始数据 RC, 从 data_categories 中取出 SN.

2) 将 RC 的文本进行分段, 对每段进行 SN 匹配, 并在匹配结果后检索数值 PP.

3) 将 [PP]-[SN] 对存入 data_collated 表中, 并将 check_status 置位为未检索.

4) 从 data_collated 读取 category_id 及 device_id 相同的数据组, 生成交叉校核表 crosstab.

5) 判断 crosstab 中数值是否相同, 若相同则将 check_status 置位为自动通过, 不同则置位为需要人工检查.

预处理完成得到的数值数据交由数据库管理员进行审核, 审核完成后的数据可通过简单检索或多字段复合检索进行查询. 同时, 管理员可随时对数据进行复核, 二次校验等操作, 以保证数据库数据的准确性.

完成审核与校验后的数据即可通过聚变实验装置数据库系统进行多维检索, 比对分析等操作.

4 总结

本文依托核安全科技领域云及其聚变物理专题数据库框架, 设计并研发了聚变实验装置数据库, 实现了聚变实验装置参数的统一入库管理, 以及多字段智能检索与多维度可视化比对分析. 此外, 收集整理了国内外典型聚变实验装置的参数信息, 并进行了入库管理与分析. 该数据库配合虚拟核电站 Virtual4DS^[15] 等核设计与安全分析系统, 可以为聚变堆系统的创新设计提供支持.

目前, 课题组正在结合聚变实验装置数据的特点, 开展数据库数据的自动爬取与自动化质量评估技术方

面的研究与功能研发工作. 同时, 除托卡马克装置外, 也正在对仿星器, 惯性约束装置等其它更多类型的聚变实验装置开展数据收集工作.

致谢

衷心感谢凤麟团队其他相关成员的帮助和指导.

参考文献

- 邱励俭. 聚变能及其应用. 北京: 清华大学出版社, 2008.
- 吴宜灿, 刘超, 宋钢, 等. 强流氦氖聚变中子源 HINEG 设计研究. 核科学与工程, 2016, 36(1): 77-83. [doi: 10.3969/j.issn.0258-0918.2016.01.011]
- 吴宜灿, 罗月童, 胡丽琴, 等. 关于核聚变研究信息化的实践与思考. 科研信息化技术与应用, 2011, 2(1): 70-77.
- Donné AJH. The European roadmap towards fusion electricity. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2019, 377(2141): 20170432. [doi: 10.1098/rsta.2017.0432]
- 高翔, 万元熙, 丁宁, 等. 可控核聚变科学技术前沿问题和进展. 中国工程科学, 2018, 20(3): 25-31.
- 王芳, 胡丽琴, 龙鹏程, 等. 聚变数据库系统 FusionDB 研发与应用. 核科学与工程, 2017, 37(1): 58-64. [doi: 10.3969/j.issn.0258-0918.2017.01.011]
- 李海峰. MVC 模式架构的应用研究. 自动化与仪器仪表, 2013, (1): 4-5, 7. [doi: 10.3969/j.issn.1001-9227.2013.01.002]
- 刘翔宇. 基于 Vue 的数据可视化系统的设计与实现 [硕士学位论文]. 北京: 北京邮电大学, 2018.
- 周晓娟, 王法能, 毛劲松. 基于 B/S 的 WEB 数据库技术. 航空计算技术, 2005, 35(4): 82-83, 87. [doi: 10.3969/j.issn.1671-654X.2005.04.024]
- Albanese R, Ambrosino G, Ariola M, et al. Overview of modelling activities for Plasma Control Upgrade in JET. Fusion Engineering and Design, 2011, 86(6-8): 1030-1033.
- 覃雄派, 王会举, 李芙蓉, 等. 数据管理技术的新格局. 软件学报, 2013, 24(2): 175-197.
- 荣艳冬. 关于 Mybatis 持久层框架的应用研究. 信息安全与技术, 2015, 6(12): 86-88.
- 沈姝. NoSQL 数据库技术及其应用研究 [硕士学位论文]. 南京: 南京信息工程大学, 2012.
- 杨武军, 张继荣, 屈军锁. 内存数据库技术综述. 西安邮电学院学报, 2005, 10(3): 95-99.
- 吴宜灿, 胡丽琴, 龙鹏程, 等. 核能信息化科研协同平台研发与虚拟核电站应用实践. 中国科学院. 中国科研信息化蓝皮书. 北京: 科学出版社, 2015. 170-179.