

联合收获机知识库数据多表联合查询方法研究

刘宏新^{1,2} 张一鸣¹ 解勇涛¹ 赵一健¹ 郭丽峰¹

(1. 东北农业大学工程学院, 哈尔滨 150030; 2. 宿迁学院机电工程学院, 宿迁 223800)

摘要: 联合收获机知识库系统使用 SQL Server 数据库,数据库中的众多数据表具有独立性,便于构建和管理。但当知识库数据量达到一定规模后,对数据表逐一查询不具有可操作性,将全部数据表融合又会导致数据结构混乱、内容表述不清、技术无法实现。针对这一问题,提出联合收获机知识库数据多表联合查询方法。从多角度划分数据表类型,分析联合收获机知识库数据存储结构,设置多表联合数据管理范围;应用 SQL(Structured query language)语言将多表信息融合为数据集并存入临时表中,实现多表联合运行;利用人机交互界面将用户查询需求转换成多表联合查询语句,生成查询结果,实现多表近似范围查询和多表精确定位查询。多表联合查询与传统单表知识查询之间的测试结果表明,多表近似范围查询比系统原有的单表近似范围查询节约用户操作时间 50% 及以上,最高达到 90.4%;多表精确定位查询比系统原有的单表精确定位查询节约用户操作时间 48.1% 及以上,最高达到 89.6%。多表联合查询的实现使联合收获机知识库系统具有实用性与可行性,可为同类知识库系统数据管理提供可借鉴的思路和方法。

关键词: 联合收获机; 知识库系统; 多表联合查询; 数据表分类; 数据同步; 人机交互

中图分类号: S225.3; TP311.1 **文献标识码:** A **文章编号:** 1000-1298(2023)05-0150-13 **OSID:**



Multi-table Joint Query Method for Combine Harvester Knowledge Base Data

LIU Hongxin^{1,2} ZHANG Yiming¹ XIE Yongtao¹ ZHAO Yijian¹ GUO Lifeng¹

(1. College of Engineering, Northeast Agricultural University, Harbin 150030, China

2. School of Mechanical and Electrical Engineering, Suqian College, Suqian 223800, China)

Abstract: The combine harvester knowledge base system uses the SQL server database, numerous data tables in the database are independent and easy to build and manage. But when the amount of knowledge base data reaches a certain size, querying data tables one by one is not actionable and merging all the data tables will lead to confusion in the data structure, unclear content expression, and technical inability to achieve. In response to this problem, a multi-table joint query method of combine harvester knowledge base data was proposed. The data table types was divided from multiple perspectives, the data storage structure of the combine harvester knowledge base was analyzed and the management scope for multi-table joint data was set. The application structured query language (SQL) fused multi-table information into a dataset and stored it into a temporary table to achieve multi-table joint operation. The human-computer interactive interface was used to convert the user query requirements into multi-table joint query statements to generate query results, and multi-table approximate range query and multi-table precise positioning query were realized. The test results between multi-table joint query and traditional single-table knowledge query showed that on the one hand, multi-table approximate range query saved user operation time by 50% or more than the original single-table approximate range query of the system, and the highest reached 90.4%; on the other hand, the multi-table precise positioning query saved 48.1% or more user operation time compared with the original single-table precise positioning query of the system, and the highest reached 89.6%. The implementation of multi-table joint query made the combine harvester knowledge base system practical and feasible and provided a reference idea and method for data management of similar knowledge base system.

Key words: combine harvester; knowledge base system; multi-table joint query; data table classification; data synchronization; human-computer interactive

0 引言

联合收获机是农业现代化发展过程中的一种典型机具^[1],我国地域差异大、农作物种类多样令联合收获机具有定制化及多样化的特征^[2],这些特征使联合收获机在设计时需因地制宜,也使联合收获机在设计过程中的限制性因素增多,再加上联合收获机本身集成度高、结构复杂的特点,最终导致其设计过程繁琐、设计周期长^[3]。建立与联合收获机相关的用于知识积累和管理平台的知识库系统,可方便设计人员灵活使用设计知识,增强知识继承性和重用性,从而提高联合收获机产品的设计效率^[4-5]。鉴于 SQL Server 数据库在易用性、可扩展性以及可靠性等方面处于世界领先地位,应用广泛^[6],机械装备知识库系统大都使用 SQL Server 数据库用于设计知识组织、存储、查询等^[7-9]。

但目前,以知识继承和设计资源重用为目的的专门知识库系统的信息查询,以及装备企业现有 PDM/PLM (Product data management / Product lifecycle management) 系统中的相关知识档案模块内容有效利用方面均存在一定的缺陷:一方面,专用知识库中不同类型的设计知识被存储于不同的数据表中,SQL Server 数据库中这种对应不同类型的单一数据表结构简单、内容明了且针对性强,用户可方便地对该数据表进行单独编辑或查询等操作,便于用户团队的分工协作和所属任务的数据管理。但是,当数据库中数据不断累积,数据表的数量达到一定规模后,用户在查询数据时为获得较全面的查询结果,需要访问更多的数据表并逐一进行查询,数据的查询就变得异常繁琐,给使用造成不便,对于大规模知识库而言,对数据表逐一定位及内容查询实际上不具有可操作性与实用性。针对这种情况,可将数据库中全部单一数据表融合成一个数据表,用户可以只针对该合成数据表进行查询,避免用户逐一访问数据表。但联合收获机知识种类繁多、信息总量大,数据表的独立性使数据需要经过分析与繁复的变换之后才能建立联系,在技术层面上该方式不具有可操作性、无法实现;即便是技术上能够实现,从结果上来看,合成数据表中的数据过于庞杂、结构混乱、内容表述不清。因此,需要在保留 SQL Server 数据库中设计知识由大量独立数据表的形式进行存储的基础上对联合收获机知识库数据多表联合查询进行研究。另一方面,目前国内外农机装备制造企业、汽车制造企业以及工程机械等行业的 PDM/PLM 系统主要以目录树结构展现系统中的全部数据集,用户可使用关键字搜索、名称搜索等搜索方式

快速定位目录树节点。对目录树节点进行查看可浏览该节点下存储的数据表内容,在该数据表中可通过使用文本筛选等方式获取具体的数据条目。企业 PDM/PLM 系统虽然数据庞大,但几乎均为档案性质的管理,对于常规个体岗位,因其对应极为有限的节点及所属数据,且数据多为简单的调用,所以并不影响正常使用。当面临以资源重用为目的的数字化与智能化设计使用需求时,因涉及相关领域大范围的数据资源作为基础,上述以有限节点权限方式进行的常规数据库结构及管理则无法满足该需求。

国内外对联合收获机知识库系统的报道较少,主要集中于系统的建立及数据的整体存储方式上,对知识库数据多表联合查询的研究未见报道,但国内外学者在其他领域中对数据的处理方式以及数据表联接方式进行了相关研究。RAHMAN 等^[10]直接将数据存储服务器与用户访问层连接,使数据中心能够在两者之间更方便的互连。BAZAGA 等^[11]在数据库中创建合成数据集,并将数据集用于多语言查询系统算法中使系统支持多个数据表执行一次查询任务。KUMAR 等^[12]为加强关系数据库管理系统与机器学习 (Machine learning, ML) 算法的集成,利用数据表之间的外键依存关系将数据表联接。LI 等^[13]利用数据表之间的共有数据将数据表联接,用户可对数据表联接后的数据集进行查询。上述 3 种对关系型数据表之间的联接为联合收获机知识库系统数据表之间的联接提供参考方案。弋伟国等^[14]将采集的果蔬质量数据加密后进行存储。程伟等^[15]标识每条遥感数据的状态信息后再将遥感数据存储到数据中心。上述两种对数据进行处理后再使用的方法为多表联合查询时数据表之间的使用开拓了思路。赵润发等^[16]用 Flink 软件作为不同类型数据的处理平台,并将 Kafka 软件用作存储数据的消息管道,该方式存储的数据在进行数据处理时反应迅速,为研究多表联合查询结果数据的存储方式提供借鉴。

综上所述,本文在前期联合收获机常规知识组织与知识库系统架构的基础上,深入开展知识库数据多表联合查询研究,解决用户无法一次性对多个数据表中数据进行查询的问题,增加联合收获机知识库系统的实用性,提高知识查询的便捷性和高效性。

1 总体技术方案

1.1 研究对象分析

研究对象(联合收获机知识库系统)为团队在项目中自主研发的支持联合收获机智能化设计的 PDM 系统的功能性子系统,具有知识管理与编辑、

浏览与查询、知识匹配及关联模型调用等功能。本文知识库及类似属性的资源型子系统在项目参与企业的实际产品研发过程中试用及验证,基本满足使用要求,但引言所述问题也显现出来。目前,联合收获机知识库系统数据表数量近 150 个,包含数千条样板数据,数据量已满足问题的充分暴露及方法检验所需的必要条目数。研究可聚焦于系统架构调整和优化,研究完成后系统可根据实际需要随时进行数据的常规扩充或大规模录入,系统功能不受数据量影响。

面对现有系统中大量独立性的数据表,系统在进行多表联合查询时需考虑哪些数据表之间可进行联合、数据表联合方式,以及联合后数据集的形式和数据查询方式。将数据表进行整理分类,利用数据表分类结果建立数据存储结构,确定可联合的数据表,并将其划分成组,每组为一个管理范围。将管理

范围中所有数据表的全部数据联合成数据集,令数据集可被统一调用即实现了多表联合运行。管理范围中的数据表在联合运行时需将各数据表中数据转换为相同的数据结构,并始终保证联合后的数据集与原数据表之间的数据保持一致。系统对数据集的查询就是系统对多个数据表中全部数据的查询,即多表联合查询。

1.2 实施方案制定

基于以上设计思路,本文提出了联合收获机知识库数据多表联合查询的研究方案,总体技术方案如图 1 所示,整个方案共包括 4 个阶段:①数据表分类:对数据表进行整理,分析数据表内容的表现形式、知识类型和数据表之间的层次关系,总结规律,将数据表进行整体分类。②数据存储结构建立及管理范围划分:根据数据表分类结果为知识库系统建立数据存储结构,令数据表能够分门别类地存储。

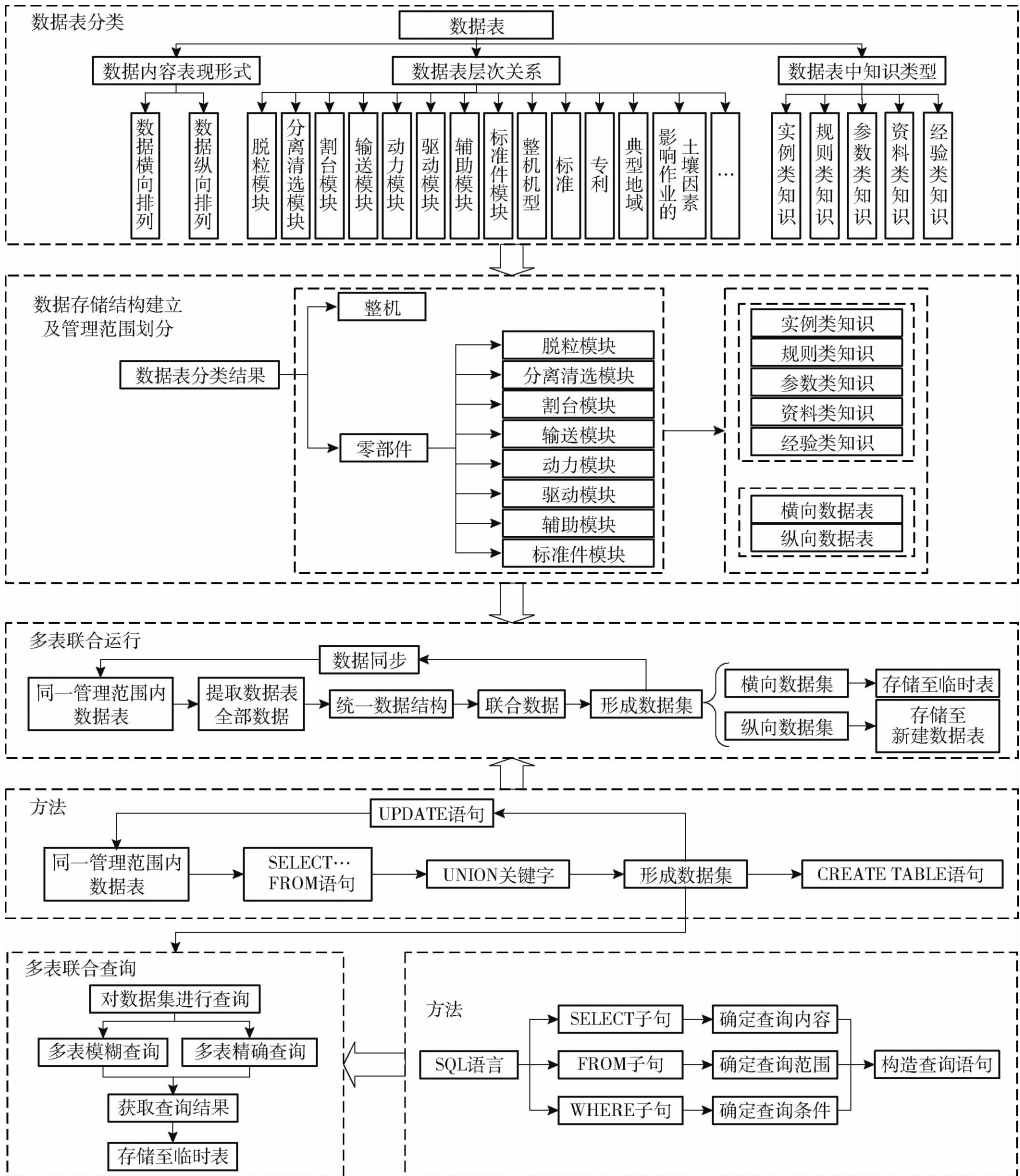


图 1 总体技术方案

Fig. 1 General technical scheme

将可融合的数据表划分成组,每组数据表即是一个管理范围,管理范围中的全部数据表在提取数据后应可以融合成数据集。③多表联合运行:根据以上管理范围划分结果,对任一管理范围的数据进行提取,并将其处理成数据集,数据集中的数据应与原数据表中的数据保持一致。形成的数据集需要为用户提供可进行数据调用、数据查询的条件。④多表联合查询:对上述数据集进行数据提取调用,用户可在数据集中查询自己的目标信息。

2 数据表在 SQL Server 数据库中的部署

2.1 数据表分析

联合收获机知识库系统将联合收获机零部件按照功能划分为脱粒模块、分离清选模块、割台模块、输送模块、动力模块、驱动模块、辅助模块和标准件模块,联合收获机零部件的相关知识也根据该划分方式进行分类并将知识进行组织整理形成各种数据表存储到 SQL Server 数据库中。其余用于描述与联合收获机相关的整机机型、标准、专利、典型地域、影响作业的土壤因素等信息按照其描述内容各自形成数据表。数据表之间的层次关系如图 2 所示。

联合收获机知识库系统在该层次关系下存储的数据表具有较小的冗余度、较高的数据独立性和易延展性,并且可在不同的用户之间实现共享。

数据表的独立性可以令联合收获机知识库系统在不影响其他数据表的情况下访问某一数据表,并对该数据表中的数据进行增加、修改、删除、查询等管理。但数据表的独立性也令系统在多表联合运行时受到限制:①联合收获机知识库系统数据表之间的关联性不强,数据表之间的数据无法关联形成完整、有联系的数据集。②数据表之间的结构各不相同,若将多个任意数据表联合起来会出现数据表结构无法融合或结构规整但内容不合理的问题。

为减少数据表的独立性对多表联合运行产生的限制,在 SQL Server 数据库的全部数据表中将同类型数据表划分到一起,形成若干个管理范围,以加强同一管理范围中数据表之间的关联性。

2.2 数据存储结构建立及管理范围划分

将数据表整体划分为联合收获机知识库系统整机类数据表和联合收获机知识库系统零部件类数据表。其中联合收获机知识库系统整机类数据表包括“整机机型”、“标准”、“专利”、“典型地域”、“影响作业的土壤因素”中的全部数据表,联合收获机知识库系统零部件类数据表包括“脱粒模块”、“分离清选模块”、“割台模块”、“输送模块”、“动力模块”、“驱动模块”、“辅助模块”和“标准件模块”中

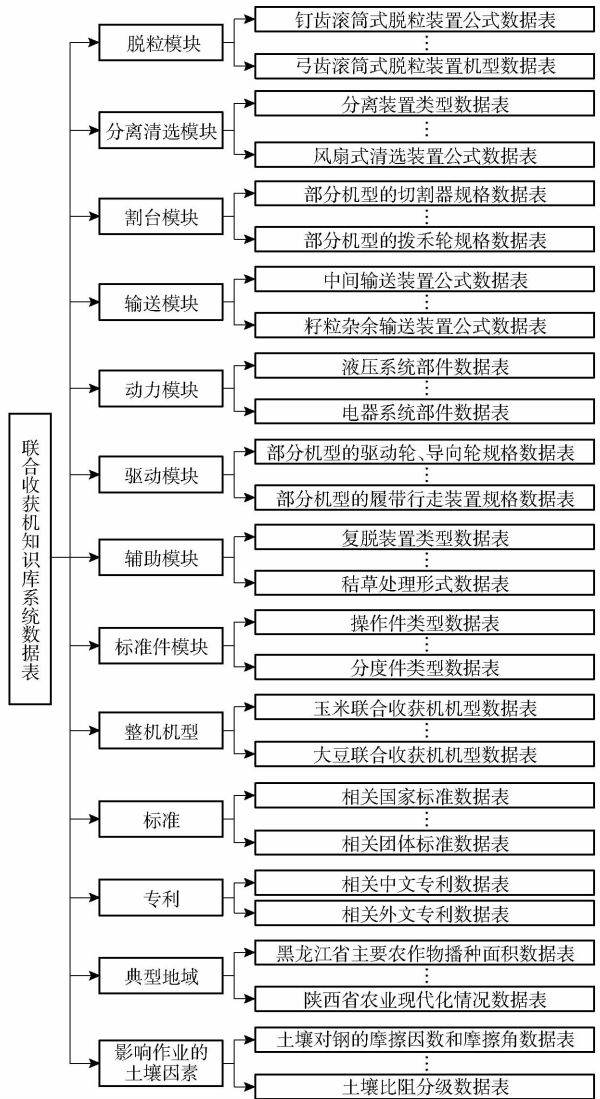


图 2 联合收获机知识库系统数据表层次关系
Fig. 2 Hierarchical diagram of data tables that belonged to harvester knowledge base system

的全部数据表。

根据联合收获机知识类型划分结果^[17]将数据表分为实例类知识数据表、规则类知识数据表、资料类知识数据表、经验类知识数据表和参数类知识数据表 5 类,同类型数据表中的数据具有相似的特点和表现形式。

实例类知识描述了联合收获机零部件的型号信息、设计产品属性、设计模型案例等信息。其中,零部件型号信息数据表中的数据纵向排列,每一列描述了该零部件的各种型号名称,列与列之间描述的是不同类型的型号名称。数据表中的每一列为一条记录,用来描述一个对象的信息,列与列之间保持独立、互不干扰,数据表中的每一行为一个字段,用来描述对象的属性。实例类知识的其他信息数据表以及其余类型知识的数据表中的数据为横向排列,其特点是数据表中的每一行为一条记录,每一列为一

个字段。横向排列的数据表和纵向排列的数据表表现形式如图 3 所示。



图 3 数据表中数据的排列形式示例

Fig. 3 Example diagrams of arrangement of data in data table

根据数据表中数据的排列情况将数据表分为数据为横向排列的数据表和数据为纵向排列的数据表(以下简称为横向数据表和纵向数据表)。数据排列情况相同的数据表具有相似的数据结构分布。

根据数据表分类结果建立联合收获机知识库数据存储结构,如图 4 所示。整个数据存储结构包含了联合收获机知识库系统的全部数据表,数据表的多类型划分使得该结构具有可扩展性,已有数据表可根据数据存储结构存入合适的位置,新增数据表可直接在该结构上添加新的节点字段进行存储。数据存储结构为数据的存储提供方便,并为多表联合运行奠定基础。

本文将可以同时进行数据存储和查询的一组数据表设置为一个管理范围,管理范围选取由同一模块层次下的部分数据表组成。在同一模块层次的基础上,选取类型和数据排列情况相同的数据表进行组织整合,共形成 54 个数据管理范围,如图 4 虚线框所示。同类型数据表之间进行联合时能够避免出现数据无法融合、数据融合后的内容不合理的情况。数据排列情况相同的数据表之间进行联合能够保证数据集结构规整。因此同一管理范围中的数据表之间结构相似,内容相关,在对其中的数据进行统一管理时能够得到结构规整、内容合理的管理结果。

3 多表联合运行实现

多表联合运行是指将任一管理范围中的全部数

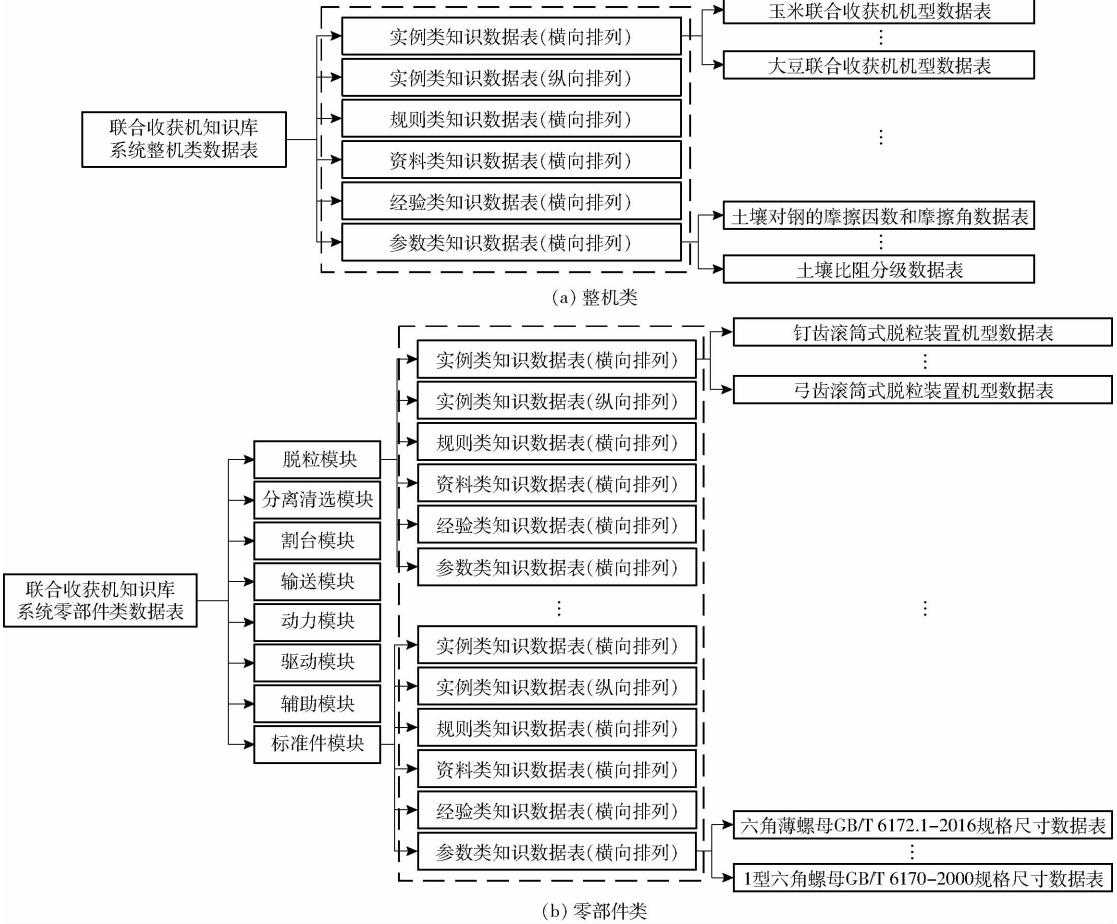


图 4 数据存储结构

Fig. 4 Stored framework of system for data

据表中的数据提取出来形成数据集,并对数据集进行调用的过程。多表联合运行主要是使用 SQL 语言实现。SQL 语言是 SQL Server 数据库的操作语言,能够独立完成 SQL Server 数据库生命周期中的全部活动,包括录入数据、建立数据库、查询、更新、维护等操作^[18-19]。

3.1 运行策略制定

实现多表联合运行首先需要提取管理范围中的全部数据表中的所有数据及结构信息,再将这些数据及结构信息融合存储到一起,达到对其统一调用的目的。横向数据表与纵向数据表的结构形式大不相同,因此横向数据表之间与纵向数据表之间的联合方式也存在差异。

横向数据表是典型的 SQL Server 数据库二维表格形式,横向数据表之间的联合是将各横向数据表中数据进行叠加的结果,形成的管理范围数据集只变动数据行数,不变动数据列数。横向数据表中的数据及结构信息的提取可使用 SQL 语言中的 SELECT 语句完成,其主要语法为^[20]

```
SELECT [ ALL/DISTINCT] <列名>
FROM <数据表名称>
其中 ALL 关键字表示显示所有行数据,
DISTINCT 关键字表示去除提取信息的重复行数据,
默认关键字为 ALL。
```

提取一个横向数据表中信息需使用一次 SELECT 语句,提取 n 个横向数据表中的数据及结构信息则需要使用 n 次 SELECT 语句。UNION 关键字能够连接各个 SELECT 语句,将各横向数据表中的数据叠加起来组成一个多表联合的数据集,该数据集包括这一管理范围内所有横向数据表中的行数据。UNION 关键字的使用方式为

```
SELECT [ ALL/DISTINCT] <列名 1>
FROM <数据表名称 1>
UNION[ ALL]
SELECT [ ALL/DISTINCT] <列名 2>
FROM <数据表名称 2>
:
```

UNION 关键字下的 ALL 关键字为可选关键字,若使用则输出全部行数据,若不使用则输出管理范围数据时将删除横向数据表之间存在的重复行数据,默认情况下是删除数据表之间的重复行。UNION 关键字在使用时,输出的管理范围数据集的结构与第 1 个 SELECT 语句下的横向数据表结构相同,因此需要将各个横向数据表的结构联合成一个能够共用的结构,使管理范围数据集输出后结构规整。

使用别名可以重新命名数据表列名,同一管理范围内的横向数据表使用相同别名,则能够实现将各横向数据表的列名统一,使各数据表拥有相同的结构,使用方法为

```
SELECT [ ALL/DISTINCT] <列名 1> ‘<别名 1>’
FROM <数据表名称 1>
UNION[ ALL]
SELECT [ ALL/DISTINCT] <列名 2> ‘<别名 2>’
FROM <数据表名称 2>
:
```

横向数据表之间的联合利用 SQL 语言实现,其形成的管理范围数据集直接存储在临时表中。临时表是临时创建的、不能永久保存的表。临时表仅对当前的用户连接可见,当用户从 SQL Server 数据库实例断开连接时,即在事务完毕或会话完毕后自动删除^[21]。即用户每次使用的数据集都是系统在用户命令发出后立即调用相关数据形成的,系统无需对数据集进行存储操作,数据集生成后自行存储于临时表中。用户对数据集使用完毕后,临时表自行删除。

纵向数据表不属于 SQL Server 数据库二维表格形式,其多表联合是将各纵向数据表中的数据逐列叠加,形成的管理范围数据集是列与列的集合,只叠加数据列数,不变动数据行数,无法使用 UNION 关键字连接各个纵向数据表。

纵向数据表之间的联合是利用新建一个数据表的方式将各纵向数据表的结构融合成新建数据表的结构,将各纵向数据表的数据叠加构成新建数据表的数据,该新建数据表即为这些纵向数据表对应管理范围的存储形式。新建数据表需先创建数据表结构,再向结构中填充数据。本文利用 CREATE TABLE 语句创建数据表结构,其语法为^[22]

```
USE <数据库名>
GO
CREATE TABLE <新建数据表名> (
<列名 1> <数据类型 1> [ NULL/NOT NULL],
<列名 2> <数据类型 2> [ NULL/NOT NULL],
...)
```

使用 USE 关键字选择数据库,在该数据库下利用 CREATE TABLE 语句创建数据表,明确新建数据表名并指出新建数据表中各列的列名、数据类型和该列是否允许为空。NULL 和 NOT NULL 关键字,决定了该列能不能出现空值,默认关键字为 NULL。

数据表结构创建好后,即可向新建数据表中填

充数据。INSERT...SELECT 语句可实现将各纵向数据表中的数据按列填充到新建数据表中,其语法为^[23]

INSERT INTO〈新建数据表名(新建数据表列名)〉

SELECT〈纵向数据表列名〉

FROM〈纵向数据表名〉

各纵向数据表中的数据是按照 SELECT 关键字后的列排序依次提取得到一个结果集,并将该结果集中的数据按照 INSERT INTO 关键字后新建数据表的列排序依次插入到新建数据表中,因此 INSERT INTO 关键字后“新建数据表列名”的排列需要与 SELECT 关键字后“纵向数据表列名”的排列相对应。

3.2 数据同步方法

纵向数据表的多表联合方式使其形成的管理范围数据集中的数据无法自动根据纵向数据表中数据的变化而变化,在多表联合运行时无法保证其输出结果的正确性和实时性。因此当用户在联合收获机知识库系统中对某纵向数据表进行知识的添加、修改或删除操作后,该纵向数据表对应的管理范围数据集需要立即进行数据的动态更新,保证管理范围数据集与纵向数据表之间的数据同步。

在 SQL Server 数据库中,对数据进行更新可通过使用 UPDATE 语句来实现,为将管理范围数据集和纵向数据表之间的数据联系起来并一一对应,在 UPDATE 语句中嵌入 SELECT...FROM...WHERE 结构和 SELECT...FROM 结构,其整体语法为

UPDATE 〈管理范围名〉

SET 〈管理范围数据集列名〉=

(SELECT 〈纵向数据表列名〉

FROM 〈纵向数据表名〉

WHERE 〈管理范围名〉. ID = 〈纵向数据表名〉

. ID)

WHERE 〈管理范围名〉. ID

IN (SELECT ID FROM 〈纵向数据表名〉)

该语法先对 SELECT...FROM...WHERE 结构进行处理,获取纵向数据表中的列数据,其中 WHERE 关键字能够令管理范围数据集和纵向数据表通过 ID 号关联,再使用“=”将管理范围数据集与纵向数据表的列一一对应。紧接着对 SELECT...FROM 结构进行处理,获得纵向数据表的各行 ID 号,并利用 IN 关键字将管理范围数据集和纵向数据表的 ID 号对应,使两者之间的行一一对应。整个 UPDATE 语句可实现根据纵向数据表数据对管理范围数据集任意列所有行数据的更新,即不论某列中的行数据变

化了多少个,管理范围数据集将更新该列的全部行数据,以防止在数据同步过程中造成错更、漏更。

但 UPDATE 语句只能对管理范围数据集的已有行数进行更新,若用户对纵向数据表进行添加数据行操作导致纵向数据表行数已经多于管理范围数据集行数,则 UPDATE 语句只更新管理范围数据集行数以内的数据,超出管理范围数据集行数的纵向数据表数据将不作为更新内容。因此在令管理范围数据集与纵向数据表数据同步时,需先保证管理范围数据集行数与纵向数据表行数一致。判断管理范围数据集行数与纵向数据表行数是否一致,可等同于判断管理范围数据集行数与各纵向数据表行数的最大值是否一致。

通过计算管理范围数据集中的各纵向数据表行数,取其最大值与管理范围数据集行数作比较,若各纵向数据表行数的最大值已超过管理范围数据集行数,则将管理范围数据集行数添加到与各纵向数据表行数的最大值一致后再对管理范围数据集进行更新,数据同步的流程如图 5 所示。

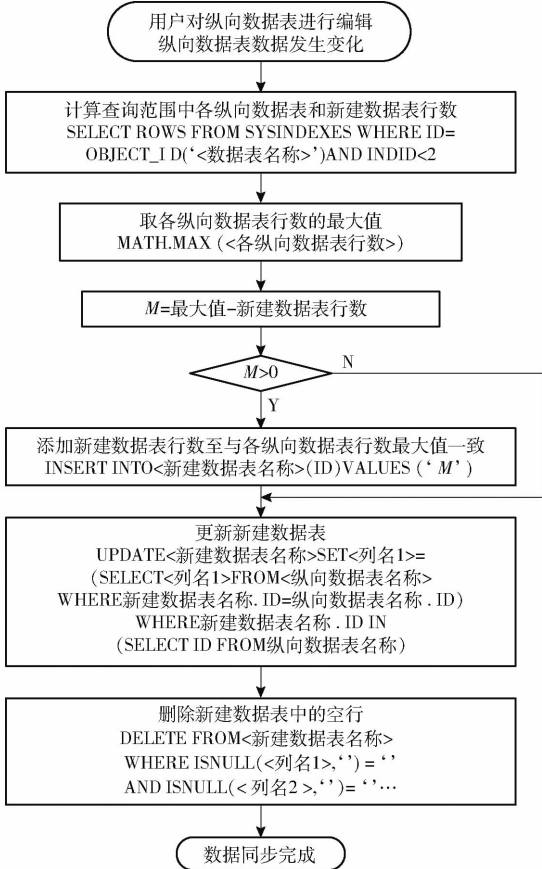


图 5 数据同步流程图

Fig. 5 Flowchart of data synchronization

4 多表联合运行应用

多表联合运行能够调用多个数据表中的数据,将这些数据进行处理后可用于用户的数据查询上,

使用户在查询时将查询范围从一个数据表扩展到一个多表联合运行管理范围数据集,实现对多个数据表数据的同时查询。多表联合查询与联合收获机知识库系统原有的知识查询相比,这种查询方式的查询范围中数据更多,获得的查询结果更全面。多表联合查询包括多表模糊查询和多表精确查询。

多表模糊查询:用户在人机交互界面上确定查询范围并通过输入字符串表明查询目标,系统在查询范围中查询包含该字符串的数据,并将数据所在数据行组成数据集作为查询结果反馈给用户。

多表精确查询:在用户非常明确自己的查询目标并对多表联合运行管理范围数据集结构有一定了解的情况下,用户通过人机交互界面将查询范围缩小至多表联合运行管理范围数据集中的某一列,以用户自身需求为查询条件,系统为用户筛选与查询条件完全匹配的数据。

4.1 查询实现原理

多表联合查询的实现实质上是用户通过在人机交互界面上进行点选或者输入对联合收获机知识库系统发出操作指令,系统将指令转换为 SQL 语言,数据库执行 SQL 语言,将得到的查询结果绑定到人机交互界面中的 DataGridView 控件上,DataGridView 控件将查询结果以表格形式显示在人机交互界面上供用户查看。本文主要使用 SQL 语言中的 SELECT 语句实现多表联合查询。SELECT 语句不但可以在数据库中模糊地查找带有某项特征的多条数据,还可以精确地查找某条数据^[24],SELECT 语句的主要语法为

```
SELECT [ ALL/DISTINCT ] <列名>  
FROM <查询范围>  
WHERE <查询条件>
```

实现多表联合查询的技术路线如图 6 所示。多表联合查询与知识库系统相同,以 Visual Studio 平台为开发工具,以 .net 为编程环境,实现多表联合查询的关键程序如图 7 所示。

4.2 查询实现方法

4.2.1 多表模糊查询

多表模糊查询的实现是通过在 SELECT 语句的 WHERE 子句中使用字符匹配符 LIKE 或 NOT LIKE 比较查询目标与查询范围中的数据是否部分匹配,将包含查询目标的数据所在行累积为数据集构成查询结果。具体语法格式为

```
WHERE <列名> [ NOT ] LIKE ‘<比较字符串>’,
```

其中,比较字符串是由查询目标字符串连同通配符一起组合而成,表 1 列出了比较常用的通配符。

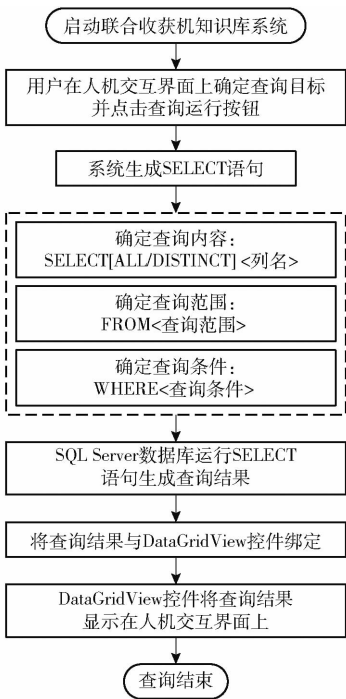


图 6 多表联合查询技术路线图
Fig. 6 Technical roadmap for multi-table joint query



图 7 多表联合查询程序框图
Fig. 7 Block diagram of multi-table joint query

表 1 常用的通配符及其含义

Tab.1 Commonly used wildcards and their meanings	
通配符	含义
%	任意多个字符
-	单个字符
[]	指定范围的单个字符
[^]	不在指定范围的单个字符

上文提到,横向数据表形成的管理范围数据集是存储在临时表中,即其在 SQL Server 数据库中并

无具体的数据表,因此在使用字符匹配符时可直接嵌入到构成横向数据表之间多表联合的语法当中,即

```
SELECT <列名> FROM <横向数据表名 1>
WHERE <列名 1> [NOT] LIKE ‘<比较字符串 1>’
UNION
SELECT <列名> FROM <横向数据表名 2>
WHERE <列名 2> [NOT] LIKE ‘<比较字符串 2>’
⋮
```

该语法既能够展现面向横向数据表的多表联合运行管理范围数据集的数据构成及结构形式,又可实现针对横向数据表的多表模糊查询,减少了语句运行过程中对各横向数据表的访问次数,缩短了命令执行时间。

由于纵向数据表形成的管理范围数据集是存储在 SQL Server 数据库中的实际存在的数据表,因此针对纵向数据表的多表模糊查询可直接在管理范围数据集中运行搭配有字符匹配符的 SELECT 语句实现,即

```
SELECT <查询范围数据集列名>
FROM <查询范围名>
WHERE<查询范围数据集列名> [NOT] LIKE
‘<比较字符串>’
```

4.2.2 多表精确查询

多表精确查询的实现是利用 SELECT 语句中 WHERE 子句的比较运算符将查询目标与数据表中的数据进行比较,筛选出存在与查询目标完全一致的数据的过程。比较运算符包括 =、<、>、<=、>=、<> 和 !=,分别表示等于、小于、大于、小于等于、大于等于、不等于和不等,其语法格式为

```
WHERE <比较表达式 1> <比较运算符> <比较表达式 2>
```

比较表达式是符号字符串、函数、计算或任何算术运算符连接的这些元素的组合,在语法中比较运算符两边的比较表达式的数据类型必须保持一致。

针对横向数据表的多表精确查询与多表模糊查询相同,直接将比较运算符嵌入到构成横向数据表多表联合的语法当中,其语法为

```
SELECT <列名> FROM <横向数据表名称 1>
WHERE <比较表达式 1> <比较运算符> <比较表达式 2>
UNION
SELECT <列名> FROM <横向数据表名称 2>
WHERE <比较表达式 3> <比较运算符> <比较表达式 4>
⋮
```

针对纵向数据表的多表精确查询是在管理范围数据集中直接运行搭配有比较运算符的 SELECT 语句实现,其语法为

```
SELECT <查询范围数据集列名>
FROM <查询范围名>
WHERE <比较表达式 1> <比较运算符> <比较表达式 2>
```

5 系统优化布局与功能测试

5.1 人机交互界面设计

登录联合收获机知识库系统后在系统主界面访问“知识浏览与查询”,并依次选择“知识查询”/“多表联合查询”,如图 8 所示,即可进入“多表联合查询”界面,该界面根据数据存储结构以及管理范围划分结果进行设计。“多表联合查询界面”包括多表近似范围查询界面和多表精准定位查询界面。其中,多表近似范围查询界面应用的是多表模糊查询方法,界面中供用户选择查询范围的目录树的设置是在数据管理范围分布结构的基础上,将树中的节点名称由完全依照数据管理范围名称这种理论性较强的表达方式转变为更直观、用户更易判断选择的表达方式。例如数据存储结构中的分离清选模块下的“规则类知识数据表”管理范围在系统多表近似范围查询界面目录树中对应的是“分离清选相关公式”节点。同时,多表精准定位查询界面应用的是多表精确查询方法,界面在数据存储结构的基础上,设置“知识模块”和“查询范围”选择框指示用户先指定查询数据所在的联合收获机所在模块,再选择该模块下的管理范围作为查询范围,设置“列名”、“条件符号”和“数值/字符串”选择框以供用户根据所选查询范围确定查询条件。

5.2 知识库系统应用方案

联合收获机知识库数据存储结构是集中式存储结构,即用户机通过给定局域网覆盖范围内的路由器连接服务器,服务器统一管理、共享全部资源并处理来自客户机的请求。

用户机无需建立知识库系统数据存储结构,可通过连接系统中心服务器局域网获取服务器中的数据存储结构。用户在使用知识库系统时调用的数据是由系统中心的服务器分享而来,系统发布的查询命令在服务器中处理后将处理结果反映在客户机上。

联合收获机知识库系统为适应实际应用需求与应用环境,具备 3 种运行方式:①对于知识库系统本身,作为一个完整系统可单独运行,用户在局域网下连接服务器的网络驱动器后,即可对该系统进行访



图 8 多表联合查询界面在系统中所处位置

Fig. 8 Position of multi-table joint query interface in system

问和应用。②对于团队自主研发的支持联合收获机智能化设计的 PDM 系统,由知识库系统、模型库系统、工程分析系统等子系统构成,各子系统独立完整、信息与数据共享,从该 PDM 系统中的特定接口可对联合收获机知识库系统进行访问,用户的请求及服务器的处理结果通过接口进行传递。③对于与企业现有 PDM 系统融合,实现资源补充与数据共享的需求情况,综合考虑数据多表联合查询方法的通用性以及企业 PDM 系统的知识产权保护和有限接口的实际情况,本团队所研究的资源型系统均采用模块化架构。因此,采用系统封装和调用的方法将数据多表联合查询方法应用于企业 PDM 系统中,实施方案如下:将知识库系统通过动态链接库技术封装,使用 Microsoft Visual Studio Installer Project 插件将知识库系统进行打包,通过程序调用直接启动知识库系统实现与企业 PDM 系统的联合运行。该方案既使知识库系统与企业 PDM 系统相对独立又将两者有机联合,充分利用了企业原有资源又赋予其新的功能,以满足产品研发工作对资源重用越来越高的要求。

5.3 多表联合数据调用

在“多表联合查询界面”中,当用户选择查询范围后,系统将自动判断该查询范围数据集是由横向数据表形成还是由纵向数据表形成。

若系统判断出用户所选择的查询范围数据集是

由横向数据表形成,系统会自行生成 SQL 语句来调用该管理范围中的全部数据表中数据并将其联合成一组数据集,随后,把数据集存放在临时表中并将其显示在界面上。当用户选择其他管理范围或者切换界面后,系统自动断开与管理范围的连接,并删除临时表。

若系统判断出用户所选择的查询范围数据集是由纵向数据表形成,系统会连接该管理范围存储在数据库中的新建数据表,并调用该数据表中数据将其显示在人机交互界面上。当用户选择其他管理范围或者切换界面后,系统自动断开与管理范围的连接,但不会删除新建数据表,该数据表依旧存储于数据库中。

5.4 多表联合查询应用实例测试

对多表联合查询功能效果进行测试,测试根据系统的运行环境于 Windows 10 操作系统上进行,测试的计算机配置如表 2 所示。

表 2 计算机配置

Tab.2 Environment parameters of computer	
名称	配置
操作系统	Windows 10,64 位
主板	FX506HM
内存	DDR4 3 200 MHz,16 GB
显卡	NVIDIA GeForceRTX 3060
CPU	11th Gen Intel(R) Core(TM) i7-11800H @ 2.50 GHz 八核
硬盘	HFM512GD3JX013N(512 GB/固态硬盘)

5.4.1 近似范围查询效果比较

用户进入“多表联合查询界面”后选择“近似范围查询”选项,获取多表近似范围查询界面。在该界面中用户首先需要在目录树上选择查询范围,系统会调用该查询范围中的全部数据。紧接着,用户在查询文本框中输入查询目标,系统会自动生成查询语句并将查询结果显示到界面上。

为检验多表近似范围查询的便捷高效性,以联合收获机知识库系统数据存储结构中的分离清选模块下的规则类知识数据表作为查询范围为例,即在描述分离清选装置公式的全部数据表中开展查询,将“高度”作为查询目标,查询与高度相关的公式。测试用户选择了解系统操作流程并可对系统进行操作的设计人员,该用户将先后使用联合收获机知识库系统原有的单表近似范围查询功能和现开发的多表近似范围查询功能分别进行查询测试。单表近似范围查询时,系统进入单表近似范围查询界面后用户需分别选择目录树中的分离装置公式、风扇式清选装置公式和风扇筛子式清选公式 3 个节点,

系统在调用其对应的数据表后用户在界面的文本框中输入查询目标字符串“高度”,系统最终将得到 3 个分散的查询结果,其中系统在节点“风扇式清选装置公式”下的查询结果如图 9 所示。从用户进入单表近似范围查询界面进行查询操作开始到查询结果全部逐一显示在交互界面上为止,该用户完成整个过程所需时间为 31 s。在执行多表近似范围查询功能时,用户只需选择多表近似范围查询界面目录树中的“分离清选相关公式”节点,并在查询文本框中输入“高度”即可得到查询结果,系统多表近似范围查询结果如图 10 所示。从用户进入多表近似范围查询界面进行查询操作开始到查询结果显示在交互界面上为止,完成整个过程所需时间为 11 s。

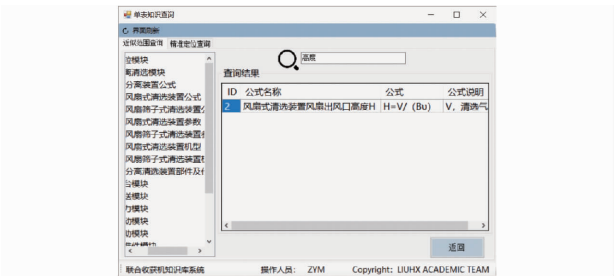


图 9 单表近似范围查询在节点“风扇式清选装置公式”下的查询结果

Fig. 9 Single table approximate range query results under node “Formula of fan type cleaning device”

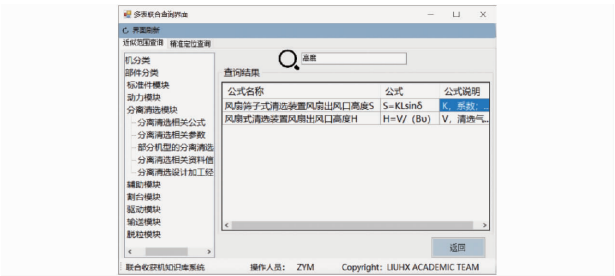


图 10 多表近似范围查询功能的查询结果

Fig. 10 Query results generated by multi-table approximate range query function

本例中对比单表近似范围查询功能需在目录树上选择 3 个节点并分别输入 3 次查询目标字符串,多表近似范围查询功能只需在目录树上选择 1 个节点并输入 1 次查询目标字符串,既简便了查询操作过程,又将查询时间缩短了将近 2/3,且一次性得到了含有“高度”字符串的全部数据。

对联合收获机知识库系统数据管理范围进行抽样,将抽样所得管理范围对应的类型数据表作为查询范围,将单表近似范围查询与多表近似范围查询的操作时间进行对比,抽样测试结果如表 3 所示。

测试结果显示多表近似范围查询比单表近似范围查询的用户操作时间减少 50% 及以上,最高节约

表 3 近似范围查询抽样测试结果

Tab. 3 Approximate range query sampling test results

查询范围	操作时间/s
脱粒模块参数类知识数据表	75/11
分离清选模块经验类知识数据表	23/11
割台模块实例类数据表	34/11
输送模块规则类数据表	22/11
动力模块资料类数据表	24/11
驱动模块实例类数据表	35/11
辅助模块资料类数据表	34/11
标准件模块资料类数据表	115/11
整机参数类数据表	34/11

注:操作时间以“单表近似范围查询测试结果/多表近似范围查询测试结果”表示,下同。

时间 90.4%。

5.4.2 精确定位查询效果比较

用户在进入“多表联合查询界面”后选择“精确定位查询”选项,获取多表精确定位查询界面。用户在界面上依次确定知识模块、查询范围后系统自动将此查询范围中的全部数据以数据表的形式呈现在界面上,并将该查询范围的所有列名放置到列名选择文本框中供用户选择要进行查询的具体列,用户再依次确定查询条件符号 >、<、=、>=、<= 以及 >=、<= 与查询范围内数据进行比较的数值或字符串后开始查询,系统即可自动生成查询语句并将满足查询条件的数据行反馈在界面上。

为检验多表精确定位查询的便捷高效性,以联合收获机知识库系统数据存储结构中的分离清选模块下的实例类知识数据表作为查询范围为例,即在描述联合收获机机型的分离清选装置参数的全部数据表中开展查询,查询机型为“4LZ-2.5”的数据。测试用户与上述进行近似范围查询测试的用户为同一人,该用户先后使用联合收获机知识库系统单表精确定位查询功能和多表精确定位查询功能开展查询测试。单表精确定位查询时,用户在知识模块“分离清选模块”下,分别选择数据表“风扇筛子式清选装置技术参数”、“风扇式清选装置技术参数”、“分离装置技术参数”作为查询范围,查询条件设置为查询机型是“4LZ-2.5”的数据,其中将“风扇筛子式清选装置技术参数”设置为查询范围时的查询结果如图 11 所示。从用户进入单表精确定位查询界面开始到完成所有查询并得到查询结果为止,该用户完成整个查询过程需执行 15 个操作步骤,用时 45 s。多表精确定位查询时,在界面上的选择文本框中分别选择“零部件分离清选模块”/“分离清选装置技术参数”/“机型”,此时条件符号这一栏的选择文本框默认为“=”且用户不可修改,接着在数

值/字符串文本框中输入“4LZ-2.5”进行查询,即可得到系统查询结果,如图 12 所示,从用户进入多表精准定位查询界面进行查询操作开始到查询结果显示在交互界面上为止,该用户完成整个过程需执行 5 个操作步骤,用时 14 s。

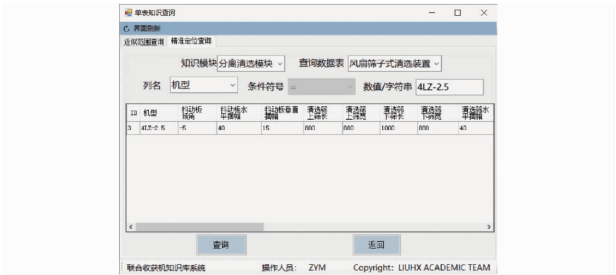


图 11 将“风扇筛子式清选装置技术参数”作为查询范围时的单表精准定位查询结果

Fig.11 Single table precise positioning query results when “Technical parameters of fan sieve type cleaning device” was used as query scope



图 12 系统中多表精准定位查询功能的查询结果

Fig.12 Query results generated by multi-table precise positioning query function in system

对比两种查询方式分析得出:多表精准定位查询的查询结果可将单表精准定位查询的结果全部涵盖,两者查询结果内容基本相同,仅在数据结构中的列名上有所不同,但并不影响查询结果的正确性。多表精准定位查询操作简便,输出的数据表结构直观明了、内容正确无误,在本例中相较于单表精准定位查询操作步骤节省时间 80.5%。

参照近似范围查询抽样测试过程,将单表精准

定位查询与多表精准定位查询的操作时间进行对比,抽样测试结果如表 4 所示。

表 4 精准定位查询抽样测试结果

Tab.4 Precise positioning query sampling test results

查询范围	操作时间/s
脱粒模块经验类知识数据表	30/14
分离清选模块实例类知识数据表	45/14
割台模块实例类数据表	44/14
输送模块规则类数据表	29/14
动力模块资料类数据表	27/14
驱动模块实例类数据表	43/14
辅助模块资料类数据表	44/14
标准件模块资料类数据表	135/14
整机资料类数据表	83/14

测试结果显示,多表精准定位查询比单表精准定位范围查询的用户操作时间减少 48.1% 及以上,最高节约时间 89.6%。

6 结论

(1)以数据表类型划分结果为基础构建的联合收获机知识库数据存储结构具有可扩展和数据管理集中的特点。其在保留单一数据表数据存储形式的同时形成多表联合数据存储形式,是研究多表联合查询方法的基础。

(2)多表联合运行将同一管理范围中所有数据表的全部数据联合成一个数据集,系统可对数据集进行调用。将管理范围作为多表联合查询功能的查询范围,使联合收获机知识库系统具有了实用性与便捷性。

(3)多表近似范围查询比系统原有的单表近似范围查询节约用户操作时间 50% 及以上,最高达到 90.4%;多表精准定位查询比系统原有的单表精准定位查询节约用户操作时间 48.1% 及以上,最高达到 89.6%。多表联合查询可满足用户快速获取所需查询数据的需求,提高了数据的查询效率。

参 考 文 献

[1] 徐立章,李洋,李耀明,等. 谷物联合收获机清选技术与装置研究进展[J]. 农业机械学报, 2019, 50(10): 1-16. XU Lizhang, LI Yang, LI Yaoming, et al. Research progress on cleaning technology and device of grain combine harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 1-16. (in Chinese)

[2] 杜岳峰,傅生辉,毛恩荣,等. 农业机械智能化设计技术发展现状与展望[J]. 农业机械学报, 2019, 50(9): 1-17. DU Yuefeng, FU Shenghui, MAO Enrong, et al. Development situation and prospects of intelligent design for agricultural machinery[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(9): 1-17. (in Chinese)

[3] 杨光,陈巧敏,夏先飞,等. 4DL-5A 型蚕豆联合收割机关键部件设计与优化[J]. 农业工程学报, 2021, 37(23): 10-18. YANG Guang, CHEN Qiaomin, XIA Xianfei, et al. Design and optimization of the key components for 4DL-5A faba bean combine harvester[J]. Transactions of the CSAE, 2021, 37(23): 10-18. (in Chinese)

[4] 刘宏新,王登宇,郭丽峰,等. 先进设计技术在农业装备研究中的应用分析[J]. 农业机械学报,2019,50(7):1-18. LIU Hongxin, WANG Dengyu, GUO Lifeng, et al. Development of advanced design technology and its application in agricultural equipment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 1-18. (in Chinese)

- [5] 刘宏新, 刘招金, 张光甫, 等. 基于 Skeleton Design 的脱粒装置交互式设计系统研究[J]. 农业机械学报, 2020, 51(12): 405–416.
LIU Hongxin, LIU Zhaojin, ZHANG Guangfu, et al. Interactive design system of threshing device based on Skeleton Design [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(12): 405–416. (in Chinese)
- [6] CHOI H, LEE S, JEONG D. Forensic recovery of SQL Server database: practical approach[J]. IEEE Access, 2021, 9: 14564–14575.
- [7] 李青林, 宋玉营, 姚成建, 等. 稻麦联合收获机清选装置智能设计与优化系统研究[J]. 农业机械学报, 2021, 52(5): 92–101.
LI Qinglin, SONG Yuying, YAO Chengjian, et al. Intelligent design and optimization system for cleaning device of rice and wheat combine harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5): 92–101. (in Chinese)
- [8] 李文斌, 李青林, 黄云林, 等. 稻麦联合收获机脱粒装置智能化设计平台构建[J]. 农业机械学报, 2020, 51(增刊2): 154–161.
LI Wenbin, LI Qinglin, HUANG Yunlin, et al. Construction of intelligent design platform for threshing device of combine harvester for rice and wheat[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(Supp. 2): 154–161. (in Chinese)
- [9] 宋正河, 毕淑琴, 金晓萍, 等. 履带式收获机械传动系快速设计推理方法[J]. 农业机械学报, 2013, 44(增刊2): 268–272.
SONG Zhenghe, BI Shuqin, JIN Xiaoping, et al. Rapid design reasoning method for crawler harvester transmission system[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(Supp. 2): 268–272. (in Chinese)
- [10] RAHMAN M N, ESMAILPOUR A. A hybrid data center architecture for big data[J]. Big Data Research, 2016, 3(C): 29–40.
- [11] BAZAGA A, GUNWANT N, MICKLEM G. Translating synthetic natural language to database queries with a polyglot deep learning framework[J]. Scientific Reports, 2021, 11(1): 1–11.
- [12] KUMAR A, BOEHM M, YANG J. Data management in machine learning: challenges, techniques, and systems [C] // Proceedings of the 2017 ACM International Conference on Management of Data, 2017: 1717–1722.
- [13] LI Xinpeng, WANG Yong, FENG Hao, et al. A parallel host log analysis approach based on spark [C] // 2018 14th International Conference on Computational Intelligence and Security (CIS). IEEE, 2018: 301–305.
- [14] 弋伟国, 何建国, 刘贵珊, 等. 区块链增强果蔬质量追溯可信度方法研究与系统实现[J]. 农业机械学报, 2022, 53(2): 309–315, 345.
YI Weiguo, HE Jianguo, LIU Guishan, et al. Development and implementation of blockchain to enhance traceability and reliability of fruit and vegetable quality [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 309–315, 345. (in Chinese)
- [15] 程伟, 钱晓明, 李世卫, 等. 时空遥感云计算平台 PIE-Engine Studio 的研究与应用[J]. 遥感学报, 2022, 26(2): 335–347.
CHENG Wei, QIAN Xiaoming, LI Shiwei, et al. Research and application of PIE-Engine Studio for spatiotemporal remote sensing cloud computing platform[J]. National Remote Sensing Bulletin, 2022, 26(2): 335–347. (in Chinese)
- [16] 赵润发, 娄渊胜, 叶枫, 等. 基于 Flink 的工业大数据平台研究与应用[J]. 计算机工程与设计, 2022, 43(3): 886–894.
ZHAO Runfa, LOU Yuansheng, YE Feng, et al. Research and application of industrial big data platform based on Flink[J]. Computer Engineering and Design, 2022, 43(3): 886–894. (in Chinese)
- [17] 刘宏新, 李金龙, 郭丽峰, 等. 联合收获机知识组织与知识库系统研究[J]. 农业机械学报, 2021, 52(2): 381–393.
LIU Hongxin, LI Jinlong, GUO Lifeng, et al. Knowledge organization and knowledge base system of combine harvester[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(2): 381–393. (in Chinese)
- [18] 李青, 钟将, 李立力, 等. 基于预训练机制的自修正复杂语义分析方法[J]. 通信学报, 2019, 40(12): 41–50.
LI Qing, ZHONG Jiang, LI Lili, et al. Self-correcting complex semantic analysis method based on pre-training mechanism [J]. Journal on Communications, 2019, 40(12): 41–50. (in Chinese)
- [19] SOEWITO B, GUNAWAN F E, HIRZI, et al. Prevention structured query language injection using regular expression and escape string[J]. Procedia Computer Science, 2018, 135: 678–687.
- [20] 陈鹤, 田秀霞, 袁培森, 等. Crypt-JDBC 模型: 洋葱加密算法的优化改进[J]. 计算机科学与探索, 2017, 11(8): 1246–1257.
CHEN He, TIAN Xiuxia, YUAN Peisen, et al. Crypt-JDBC model: optimization of onion encryption algorithm[J]. Journal of Frontiers of Computer Science and Technology, 2017, 11(8): 1246–1257. (in Chinese)
- [21] 白梅, 王京徽, 王习特, 等. PSP: 一种高效的偏序域上 skyline 查询处理方法[J]. 湖南大学学报(自然科学版), 2020, 47(8): 9–20.
BAI Mei, WANG Jinghui, WANG Xite, et al. PSP: an efficient skyline computation method for partially ordered domains[J]. Journal of Hunan University(Natural Sciences), 2020, 47(8): 9–20. (in Chinese)
- [22] GUO D, ONSTEIN E. State-of-the-art geospatial information processing in NoSQL databases[J]. ISPRS International Journal of Geo-Information, 2020, 9(5): 331–351.
- [23] SCHREINER G A, DUARTE D, DOS SANTOS M R. Bringing SQL databases to key-based NoSQL databases: a canonical approach[J]. Computing, 2020, 102(1): 221–246.
- [24] 张晨煜, 刘文洁, 庞天泽, 等. 基于分布式数据库的相关子查询优化[J]. 西北工业大学学报, 2021, 39(4): 909–918.
ZHANG Chenyu, LIU Wenjie, PANG Tianze, et al. Optimization of correlate subquery based on distributed database[J]. Journal of Northwestern Polytechnical University, 2021, 39(4): 909–918. (in Chinese)