文章编号：2096-1472(2020)-05-31-04
DOI:10.19644/j.cnki.issn2096-1472.2020.05.009

两种非匀质Excel表转换成关系数据库表的方法

方木云，赵长鲜，张祝梦

(安徽工业大学计算机科学与技术学院，安徽 马鞍山 243002)
✉fangmy@ahut.edu.cn; 1826939647@qq.com; 2602617221@qq.com

摘　要：依据表中数据的特点，Excel表可以分为匀质和非匀质两种类型，而关系数据库表只有匀质一种类型，所以非匀质的Excel表数据不能直接导入到关系数据库表中，需要进行表结构的匀质化转换。为了解决这一问题，提出了两种非匀质EXCEL表转换成关系数据库表的方法，实现了非匀质Excel表结构向关系数据库表结构的转换，并用C＃编程实现了Excel表数据向关系数据库表的自动导入。

关键词：匀质Excel；非匀质Excel；关系数据库
中图分类号：TP311　**文献标识码**：A

Two Methods of Transforming Inhomogeneous Excel Table into Relational Database Table

FANG Muyun, ZHAO Changxian, ZHANG Zhumeng

(*School of Computer Science and Technology*, *Anhui University of Technology*, *Ma'anshan* 243002, *China*)
✉fangmy@ahut.edu.cn; 1826939647@qq.com; 2602617221@qq.com

Abstract: According to the characteristics of the data in the table, the Excel table can be divided into homogeneous and inhomogeneous types, while the relational database table has only one type, so the inhomogeneous Excel table data cannot be directly imported into the relational database table, and it is required to implement the homogenization transformation of the table structure. In order to solve this problem, the paper proposes two methods of transforming inhomogeneous Excel table into relational database table. The study achieve the transformation from inhomogeneous Excel table structure to relational database table structure, and the automatic import of Excel table data into relational database table in C# programming language.

Keywords: homogeneous excel; inhomogeneous excel; relational database

1 引言(Introduction)

Excel广泛应用在日常办公的数据处理中。Excel的一个突出特点是采用表格方式管理数据，所有的数据和信息都以工作表的二维表格形式管理，单元格中数据间的相互关系一目了然。很多信息系统早期都是使用Excel来进行管理，不少单位的财务系统甚至到现在还在使用Excel。随着可视化编程语言和关系数据库的出现，很多应用系统开始向C/S（Client/Server)和B/S（Browser/Server)模式的信息管理系统迁移，Excel系统逐渐被替代。尽管Excel表跟关系数据库表一样采用二维表管理数据，可是很多Excel表数据不能行列对应地转换成关系数据库表数据。如何将单位内部已经使用多年的各种Excel表结构快速有效地转化成关系数据库的表结构并实现数据自动导入是一个重要的应用问题[1−9]。

关系数据库表只提供数据标准化管理，不提供自由编辑，其列数据类型约束强，一列只能全部填写数字或者全部填写文字；而Excel对列数据类型不强制约束，一列可以填写数字、文字或为空。所以用户经常随心所欲地使用Excel表，这种用户友好性导致表格数据多种多样。

依据所存数据的特点，Excel表可以分为匀质和非匀质两种类型。匀质的Excel表结构是指列的数据类型一致和列的数

基金项目：安徽省高校自然科学研究重大项目“基于互联网+ERP建设离散制造业智慧工厂”(KJ2018ZD007).

据行相等，如学生表、课程表、选课表等；非匀质的Excel表结构是指列的数据类型不一致或列的数据行不相等。匀质的Excel表结构可以行列对应地直接转换成关系数据库表结构；非匀质的Excel表结构不能直接转换成关系数据库表结构，所以需要进行匀质化转换。

在实际开发中，由于缺乏经验，很多开发者不区分匀质和非匀质Excel，直接将所有Excel表行列对应地直接映射到关系数据库表。结果出现不少表格的数据不便于检索和无法扩展，用户数据需求一旦增加，软件就无法使用，导致开发出来的系统是“僵尸”系统。所以需要提出一种非匀质Excel表转换成关系数据库表的方法。文献[1]—文献[7]开发了将Excel数据导入数据库的工具，没有区分不同的Excel，文献[8]—文献[9]考虑了Excel的不规则性，从而开发了一个导入工具。本文将Excel的不规则性区分为行非匀质和列非匀质的两种Excel。

2 非匀质Excel表的定义(Definition of non-homogeneous Excel tables)

关系数据库表是强约束的，有如下特征：(1)一个表描述的是一个实体或者实体之间的一个联系；(2)一个列描述一个属性，也称为字段，是存储和检索数据的关键；(3)每个列的数据类型是唯一的，如：只能是数值或者是文本，不能同时存数值和文本；(4)表的行数是相等的，也就是每个字段有相同的行值，允许有缺省值。Excel表是弱约束的，有如下特征：(1)一个表可以记录任意一个实体和联系；(2)一个单元格是存储和检索数据的关键；(3)一个列可以记录不同类型的数据；(4)表的行可以长短不一致。

Excel表通过单元格进行数据的相对和绝对引用，通过VLOOK来查找和调用数据。关系数据库表通过列使用SELECT结合WHERE条件来查找和调用数据。

完全不同于关系数据库一个表只能记录一个实体或联系的数据，Excel表弱约束带来的宽松和灵活性使得其得到广泛应用，可以同时用来记录结构化和非结构化的数据，可以记录多个实体和联系，导致表格可以分为匀质和非匀质两种类型。

定义：当Excel表存储结构化数据时，完全按照关系数据库表模式来记录数据的Excel表称为匀质Excel表；凡是不按照关系数据库表模式来记录数据的Excel表称为非匀质Excel表。其中，数据行不相等的Excel表称为行非匀质Excel表；数据列类型不一致的Excel表称为列非匀质Excel表。

结论：当Excel表存储结构化数据时，所有非匀质Excel表都可以归入行非匀质Excel表或者列非匀质Excel表，不存在行列同时不匀质的Excel表。

证明：一个正确表达结构化信息的Excel表是二维表，利用某一列或某一行来存储某个属性，其不匀质只能来自列或行，不能同时来自列和行，否则无法正确表达信息。

下面举出几个实例(如表1—表3所示)，分别是匀质Excel表、行非匀质Excel表、列非匀质Excel表。

表1 匀质Excel表

Tab.1 Homogeneous excel table

序号	A	B	C	D	E
1	学号	姓名	性别	年龄	籍贯
2	201901	王鹏	男	22	安徽
3	201902	张萍	女	22	江苏

表2 行非匀质Excel表

Tab.2 Line non-homogeneous excel table

序号	A	B	C	D	E	F
1	建设投资成本	政府付费收入	使用者付费收入	运营维护成本	税金附加项目	股权出资单位
2	建筑工程费用	建设投资本金回收			城市建设维护费	安阳文化影视出版集团
3	安装工程费用	建设投资资金收益			教育费附加	上海宝冶集团有限公司
4	设备购置费用				地方教育费附加	中冶建信基金管理(北京有限公司)
5	征地拆迁费用					中奥广场管理集团有限公司
6	建设其他费用					机械工业第六设计研究院有限公司
7	工程预备费用					

表3 列非匀质Excel表

Tab.3 Column non-homogeneous excel table

序号	A	B	C
1	学号	201901	201902
2	姓名	王鹏	张萍
3	性别	男	女
4	年龄	22	22
5	籍贯	安徽	江苏

表1是匀质Excel表，其表结构和数据可以行列对应地直接导入到关系数据库表当中；表2是行非匀质Excel表，行的长度不一致；表3是列非匀质Excel表，列的属性不一致，列B既存了姓名这种文本型数据又存了年龄这种数字型数据。表2和表3的数据结构和值都不能行列对应地直接导入到关系数据库表当中，需要进行匀质化转换。

下面讨论如何将行非匀质和列非匀质Excel表的数据结构转换为关系数据库表结构，并如何编程将对应数据导入到关系数据库表中。

3 非匀质Excel表向数据库表的转换方法(Transformation method of non-uniform excel table to database table)

3.1 行非匀质的Excel转换方法

针对表2这种行非匀质的Excel表，首先进行形式化描述：将Excel表中数据划分为列字段的名称和列字段的值两个部分。Excel表中第一行数据为列字段的名称，分别标识为F1，F2，...，Fn，其中n>1；Excel表中剩下的数据为列字

段的值，分别标识为：

VF11，VF12，...，VF1x；

VF21，VF22，...，VF2y；

...

VFn1，VFn2，...，VFnz，

其中，x>=1，y>=1，z>=1。

形式化描述的结果如表4：

表4 行非匀质Excel表

Tab.4 Line non-homogeneous excel table

序号	A	B	C	D
1	F1	F2	---	Fn
2	VF11	VF21	---	VFn1
3				
4				
5	VF12	VF22	---	VFn2
6	---	---	---	---
7	---	---	---	VFnz
8	VF1x	---	---	
9		VF2y	---	

然后在关系数据库表中建立类别字段TypeField，类别字段TypeField用于存放列字段的名称标识：F1，F2，...，Fn。

最后在关系数据库表中再建立内容字段ContentField，内容字段ContentField用于存放列字段的值标识：

VF11，VF12，...，VF1x；

VF21，VF22，...，VF2y；

...

VFn1，VFn2，...，VFnz。

转换后的结果如表5所示。

表5 行非匀质Excel转换成的关系数据库表

Tab.5 Line non-homogeneous excel table translating into relation database table

序号	A	B
1	TypeField	ContentField
2	F1	VF11
3	---	---
4	F1	VF1x
5	F2	VF21
6	---	---
7	F2	VF2y
8	---	---
9	Fn	VFn1
10	---	---
11	Fn	VFnz

经过这样的转换后，数据存两列，整个变匀质了。以表2为例，第一行的【建设投资成本，------，股权出资单位】成为类别字段的值，【建设工程费用，------，工程预备费用】随着类别【建设投资成本】循环存，这样数据就可以扩充了。

3.2 列非匀质的Excel转换方法

针对表3这种列非匀质的Excel表，首先进行形式化描述：将Excel表划分为列字段的名称和列字段的值两个部分，Excel表中第一列数据为列字段的名称，分别标识为F1，F2，...，Fn，其中n>1；Excel表中剩余数据为列字段的值，分别标识为：

VF11，VF12，...，VF1x；

VF21，VF22，...，VF2x；

...

VFn1，VFn2，...，VFnx，

其中，x>=1。

形式化描述的结果如表6所示。

表6 列非匀质的Excel表

Tab.6 Column non-homogeneous excel table

序号	A	B	C	D	E
1	F1	VF11	VF12	---	VF1x
2	F2	VF21	VF22	---	VF2x
3	---	---	---	---	---
4	Fn	VFn1	VFn2	---	VFnx

然后直接把F1，F2，...，Fn建成关系数据库表中的字段。

最后导入其对应的字段值：

VF11，VF12，...，VF1x；

VF21，VF22，...，VF2x；

...

VFn1，VFn2，...，VFnx。

转换的结果如表7所示。

表7 列非匀质Excel转换成的关系数据库表

Tab.7 Column non-homogeneous excel table translating into relation database table

序号	A	B	C	D
1	F1	F2	---	Fn
2	VF11	VF11	---	VF11
3	VF12	VF12	---	VF12
4	---	---	---	---
5	VF1x	VF1x	---	VF1x

经过这样的转换后，整个表变匀质了，这样数据就可以扩充了。通过SELECT结合WHERE条件来查找数据。列非匀质Excel表比行非匀质Excel表容易转换，转换前也容易理

解，转换时不需要另外建立字段名。行非匀质Excel表“欺骗性”比较强，没有经验的开发者，可能不进行转换就直接行列对应地转换成数据库表，这样的系统没有扩展性，使用不方便。

4 编程实现自动转换结构和导入数据(Programming to achieve automatic transformation structure and import data)

提出了行非匀质Excel表和列非匀质Excel表向关系数据库表转换的方法之后，可以编程自动实现结构的转换和数据的导入。下面分别给出两个结构转换和数据导入算法：

(1)行非匀质Excel表的转换算法

Step1：利用Create table命令在数据库中建立具有两个字段TypeField和ContentField的表T1；

Step2：读取Excel表，以F1至Fn作为外循环，以VF11至VF1x等作为内循环，将值循环填入到TypeField和ContentField两个字段下面；

Step3：循环终止。

(2)列非匀质Excel表的转换算法

Step1：读取Excel表，以F1至Fn作为循环拼接字符串，利用Create table命令在数据库中建立具有n个字段F1，...，Fn的表T1；

Step2：读取Excel表，以F1，...，Fn作为读取条件，循环将对应的值填入到其字段下面；

Step3：循环终止。

在PPP财务评价软件中，利用C＃编程，将两种非匀质的Excel表成功的转换到关系数据库当中。下面以行非匀质的Excel表为例，针对表2的Excel表转换结果如下：关系数据库表结构如图1所示，关系数据库表记录如图2所示。

LAPTOP-UKJMVO8I...o.TableFromExcel

列名	数据类型	允许 Null 值
TypeField	nvarchar(50)	☑
ContentField	nvarchar(50)	☑

图1 关系数据库表结构

Fig.1 Relational database table structure

	TypeField	ContentField
	股权出资单位	安阳文化影视出版集团
	股权出资单位	上海宝冶集团有限公司
	股权出资单位	机械工业第六设计研究院有限公司
	股权出资单位	中奥广场管理集团有限公司
	股权出资单位	中冶建信基金管理（北京）有限公司
	建设投资成本	建筑工程费用
	建设投资成本	工程预备费用
	建设投资成本	设备购置费用
	建设投资成本	建设其他费用
	建设投资成本	安装工程费用
	建设投资成本	征地拆迁费用
	使用者付费收入	使用者付费1
	使用者付费收入	使用者付费2
	税金附加项目	地方教育费附加
	税金附加项目	城市建设维护费
	税金附加项目	教育费附加
	运营维护成本	运营维护成本1
	运营维护成本	运营维护成本2
	政府付费收入	建设投资本金回收
▶	政府付费收入	建设投资资金收益

图2 关系数据库表记录

Fig.2 Relational database table records

5 结论(Conclusion)

在实际应用当中，非匀质Excel表存在的原因有以下几点：(1)数据的直观性，表达方式比较直观，便于使用者瞬间理解；(2)数据量不大，不需要扩充的时候，通过VLOOK来加工查找数据方便；(3)数据的原始生成者，没有想到数据扩充的问题，因为Excel是按照文件来扩充数据的。

非匀质Excel表转向关系数据库的时候，不采用类似本文提出的转换方法，如果都是行列对应直接导入，导致数据库不能扩充，所作出来的软件将会成为“僵尸”系统。看起来很好，但无法扩充。

将非匀质Excel表分为行非匀质和列非匀质两种类型，并提出结构转换和数据自动导入的方法，并利用C＃编程实现，成功应用到PPP财务评价软件系统的项目开发当中。

参考文献(References)

[1] 陈小龙,陈绮璟.基于C#.NET实现Excel数据导入数据库技术[J].计算机与网络,2019,45(23):46-47.

[2] 周晓俊.ASP.NET中Excel数据处理的技术实现[J].信息与电脑(理论版),2018(06):113-115.

[3] 罗雅丽.关于大数据导入数据库的方法探究[J].电脑编程技巧与维护,2019(08):111-113.

[4] 魏景东.将Excel表数据导入MS SQL Server数据库表的一种有效方法[J].电脑编程技巧与维护,2013(07):53-56.

[5] 罗丽云,段艳萍,简碧园.ASP.NET中导入Excel数据到数据库的应用与实现[J].科技创新与应用,2015(29):89.

[6] 詹重咏.MySQL数据库中数据导入与导出探析[J].数字技术与应用,2017(12):231;233.

[7] 陈道远,孙兆辉.基于XML配置的Excel通用导入组件设计与应用[J].电脑编程技巧与维护,2019(08):99-100.

[8] 武彤,陆昱霖.基于XML映射模板实现不规则Excel数据的转换[J].计算机技术与发展,2015,25(07):209-212.

[9] 张琦.基于.Net技术实现Excel数据抽取及批量入库[J].电脑编程技巧与维护,2018(09):85-88.

作者简介：

方木云(1968-)，男，博士，教授.研究领域：软件工程，信息系统.

赵长鲜(1997-)，女，硕士生.研究领域：软件工程，信息系统.

张祝梦(1998-)，女，本科生.研究领域：软件工程，信息系统.