

基于 VB 的 SCI 期刊 Excel 数据检索及影响因子提取^①

马 超, 张显库, 杨光平, 张志恒, 冯永孝

(大连海事大学 航海动态仿真和控制实验室, 大连 116026)

摘 要: 为了方便科研工作者了解本领域 SCI(Science Citation Index)期刊的影响因子, 编程实现了针对每年公布的 11000 余种 SCI 期刊某一领域数据的筛选, 以得到与本学科相关期刊的影响因子情况. SCI 收录的期刊涉及近百个研究领域, 而每个研究者只关心其中的一个或两个研究领域. 目前, 科研工作者多利用 Excel 工具, 人工手动处理这些期刊数据, 其工作量巨大且准确性不能得到保障. 用 VB 6.0(Visual Basic6.0)编程实现对 SCI 期刊数据库工作簿的批量数据处理, 其过程快速、便捷、准确率高且适用范围广.

关键词: VB; 批量处理数据; Excel 工作簿; SCI 数据库; 正选反删

Retrieval of Excel Data and Impact Factor From SCI Journals Based on VB

MA Chao, ZHANG Xian-Ku, YANG Guang-Ping, ZHANG Zhi-Heng, FENG Yong-Xiao

(Key Laboratory of Marine Simulation and Control, Dalian Maritime University, Dalian 116026, China)

Abstract: In order to facilitate the scientific research workers to know the impact factor of SCI (Science Citation Index) journal in their own field, the program realizes a filter of more than 11,000 kinds of SCI journals in a certain field that are published annually to get impact factors of the journal that are related to their own subject. SCI journal involves nearly 100 research fields, but each researcher only cares about one or two research areas. At present, the scientific research workers mostly use Excel tool to handle the journal data by hand. Thus it needs a huge workload and the accuracy cannot be guaranteed. The batch data of workbook in SCI journal dataset can be fast processed by VB6.0 (Visual Basic6.0) program, and the procession has the characteristics of high accuracy and high-speed has a wide range of application.

Key words: Visual Basic(VB); batch processing data; excel workbooks; SCI database; postive select and anti-delete

美国科学情报研究所出版的《科学引文索引》(Science Citation Index, 简称 SCI), 是世界著名的期刊文献检索工具, 其收录的文献能够全面覆盖全世界最重要和最有影响力的研究成果^[1]. 这些宝贵的资料是后续科学研究的参考, 对于科研人员有着至关重要的价值. 也正是 SCI 数据库内容繁多, 对于每名科研人员的学习和研究都十分不方便; 尤其是每年 SCI 新收录的论文对于科研至关重要, 但往往对每个学科又是无用的信息居多. 目前, 多数科研人员都是用 Office 软件 Excel 处理 SCI 数据库, 以获得所需的内容. 可是, 这样处理 SCI 数据库不仅工作量大, 还容易因为人的疲劳而产生误删除, 影响结果的准备性. 所以急需寻

找一种简便易行的方法处理新版 SCI 收录的数据内容, 用 VB 编程实现 Excel 恰好能很好地解决这个问题.

VB6.0(Visual Basic6.0)是一种由微软公司开发的结构化的、模块化的、面向对象的、包含协助开发环境的事件驱动为机制的可视化程序设计语言, 是设计应用程序强有力的开发工具^[2]. Excel 是微软办公套装软件的一个重要的组成部分, 它可以进行各种数据的处理、统计分析和辅助决策操作, 广泛地应用于管理、统计财经、金融等众多领域. 由于 Excel 的应用程序对象是外部可以创建的对象, 所以可以从 VB 应用程序内部来程序化操纵 Excel, 用 VB 应用程序来调用指定的 Excel 文件—SCI 数据库工作簿, 并在其内

① 基金项目:国家自然科学基金(51109020,51409033);国家高技术研究发展计划(863)(2015AA016404);中央高校基本科研业务费专项资金(3132014302)

收稿时间:2016-09-20;收到修改稿时间:2016-10-24 [doi: 10.15888/j.cnki.csa.005785]

部来进行界面化程序编辑,以达到对 SCI 数据库工作簿进行程序化数据处理的目的. 本文结合作者自己编写的 VB 程序,谈谈如何利用 VB 编程实现对含大量数据的 SCI 数据库工作簿进行批量数据处理.

1 Excel常用对象及VB调用Excel

1.1 Excel 常用对象

为了在 VB 应用程序中调用 Excel, 必须要了解 Excel 对象模型. Excel 对象模型包括了 128 个不同的对象, 本文涉及 VB 应用程序调用 Excel 的对象包括以下几种^[3,4]:

(1) Application 对象

Application 对象是 Excel 对象模型的顶层, 表示整个 Excel 应用程序.

(2) Workbook 对象

Workbook 对象直接隶属于 Application 对象的底层, 表示一个 Excel 工作簿文件. Workbook 对象的集合则构成 Workbooks 对象, 可以通过 Workbooks 集合或表示当前活动工作簿的 Active Workbook 对象访问 Workbook 对象.

(3) Worksheet 对象

Worksheet 对象包含于 Workbook 对象中, 表示一个 Excel 工作表. 可以通过 Worksheets 集合来访问、激活、增加、删除、更名工作表.

(4) Range 对象

Range 对象是 Excel 应用程序中最常用的对象, 代表工作表的某一单元格或多个单元格、某一选定区域等.

(5) Cells 对象

Cells 对象包含于 Worksheet 对象中, 表示 Excel 工作表中的一个单元格.

1.2 VB 应用程序调用 Excel

在 VB 应用程序中调用 Excel, 实质是将 Excel 作为一个外部对象来引用, 由 Excel 对象模型提供能从 VB 应用程序内部来程序化操纵的对象以及相关的属性、方法和事件^[5]. 为了能从 VB 应用程序中访问 Excel 丰富的内部资源, 使 Excel 应用程序运行得更快, 需要在 VB 工程中添加对 Excel 类型库的引用. 具体步骤如下^[6]:

(1) 打开 VB6.0, 在“工程”菜单中选择“引用”.

(2) 在“引用”对话框中选择 Excel 类型库:

"Microsoft Excel 16.0 Object Library". 由于 Excel 版本不同, 所以这个选项的版本号不同^[7].

(3) 单击左边小方框, 使之出现“√”符号.

(4) 按“确定”退出.

2 VB处理SCI数据库工作簿实例应用

2.1 数据背景

本程序待处理文件是 SCI 数据库工作簿, 其内容包括期刊名称(B 列)及影响因子(C 列), 共 11992 行, 除标题结尾外, 有效数据 11990 行. 作者需要从含有 11990 行不同学科数据的 SCI 数据库工作簿中获取与交通运输工程学科相关的内容.

2.2 程序界面设计

程序界面设计如图 1. 界面左侧为可编辑区域, 使用者可填写想要保留的学科数据含有的关键词, 程序设计时默认的 16 个关键词是与交通运输工程学科相关的常见词, 包括 traffic, ocean, ship, isa, naving, transport, model, automat, control, naval, simulat, neural network, marine, information, fuzzy, cybernet; 界面右侧是以关键字为基础而设置的可选的待删除类别, 包括医学类, 物理类, 生物类, 化学类, 人文类, 经济类, 环境类, 农林及天文类. 其中, 关键词的选取是以 2014 年及 2015 年 SCI 期刊数据库为基础, 用数学统计方式计算得出.



图1 程序界面图

2.3 程序原理及部分源代码

2.3.1 引用 Excel 对象

在 VB 应用程序中调用 Excel, 就是使用 Application 对象的方法、属性和事件^[8,9]. 为此, 首先应声明对象变量:

Dim xlapp As Object

在声明对象变量后, 可以使用 CreateObject 方法引用已存在的 Application 对象^[10]:

```
Set xlapp = CreateObject("Excel.Application")
```

其次对工作簿声明对象变量:

Dim xlbook As Object

然后按路径调取已存在的 Workbooks

```
Set xlbook = xlapp.Workbooks.Open(path)
```

本程序文件路径采取的是 InputBox 语句方法进行输入, 以适用不同情况, 代码如下:

```
path=InputBox("请输入待处理的 SCI 数据库工作簿路径","路径","D:\DLMU\LNG310\2016-SCI")
```

2.3.2 程序运行原理

程序运行原理流程如图 2 所示. 程序处理 SCI 数据库工作簿提供了“正选”, “反删”, 以及两者混合应用 3 种方式. 正选, 即通过输入本专业期刊名字中常出现的词汇, 选取保留 SCI 数据库工作簿中含有这些关键词的行, 删除其余无关行. 反删, 即在不能确定本专业期刊名字常用词汇或不能穷尽这些词汇的情况下, 通过本程序内已定义的分类学科, 直接删除无关学科期刊. 两者混合应用, 是既通过对输入的需要学科期刊名字中常用词汇进行保留, 又删除无关学科的处理方式, 来得到理想的数据.

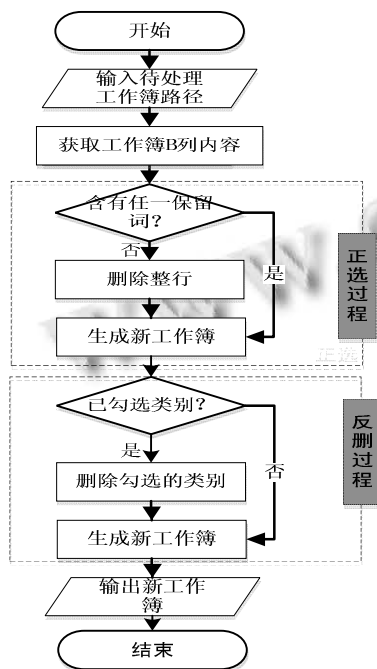


图 2 程序运行流程图

程序功能的实现关键是 Instr()函数的应用. Instr()函数语法如下:

```
InStr([start, ]string1,string2[,compare])
```

Start, 可选参数. 为数值表达式, 设置每次搜索的起点. 如果省略, 将从第一个字符的位置开始.

String1, 必要参数. 接受搜索的字符串表达式.

String2, 必要参数. 被搜索的字符串表达式.

Compare, 可选参数. 指定字符串比较方式.

Instr()函数的返回值为, 从 Start 位置开始, String2 在 String1 中的起始位置. 如果没有找到, 则返回 0. 本程序正是通过 Instr()函数返回值, 来确定 SCI 数据库工作簿的期刊名字列的单元格是否存在关键字. 函数返回值不为 0, 则说明单元格存在关键字; 函数返回值为 0, 则说明单元格不存在关键字.

程序设计如下:

1) 正选: 当程序界面左侧任意文本框填有关键词时, 右侧复选框未勾选的情况下, 程序会检索并保留含有任一关键词的行. 利用 Instr()函数判断 SCI 期刊数据库工作簿内容是否包含文本框内容, 并获得对应返回值, 分别为 j_1, j_2, \dots, j_{16} , 并令:

$$J = \sum_{i=1}^{16} j_i \quad (1)$$

判断 J 是否为 0: J 不为 0, 说明 SCI 期刊数据库工作簿的该行内容包含文本框内容, 该行予以保留; J 为 0, 说明 SCI 期刊数据库工作簿的该行内容不包含文本框内容, 该行予以删除.

程序部分重要代码如下:

```
strx = xlsheet.Cells(i, 2) '注释: 获取 i 行 B 列单元格内容
```

```
If Text1 = "" Then j1 = 0 Else j1 = Val(InStr(1, strx, t1, 1))
```

.....

```
j = j1 + j2 + j3 + j4 + j5 + j6 + j7 + j8 + j9 + j10 + j11 + j12 + j13 + j14 + j15 + j16
```

```
If j = 0 Then
```

```
xlsheet.Rows(CStr(ii) & ":" & CStr(ii)).Delete '注释: 不含有任一关键词则删除整行
```

```
End If
```

2) 反删: 当程序界面左侧文本框没有填写, 右侧勾选任意复选框时, 程序直接运行界面右侧删除类别的命令. 此时程序实现的即是“反删”过程. 该过程的

关键在于识别左侧关键词是否填写,只有左侧关键词全未填写时,程序才可以实现“反删”,否则程序实现的是“正选”“反删”混合应用。程序通过判断左侧文本框内是否为空串来识别左侧关键词是否填写。其部分代码如下:

```
If Text1 = "" And Text2 = "" And Text3 = "" And  
Text4 = "" And Text5 = "" And Text6 = "" _  
And Text7 = "" And Text8 = "" And Text9 = "" And  
Text10 = "" And Text11 = "" And Text12 = ""  
And Text13 = "" And Text14 = "" And Text15 = ""  
And Text16 = "" Then
```

j=1'注释:左侧为空,强制返回值为1,正选过程失效

3) 混合应用:当程序界面的左侧文本框填有需要保留的关键词,并在界面右侧的复选框勾选需要删除的类别时,程序首先会自动筛选保留含有界面左侧关键词的行,然后根据软件内已有的各个类别的关键词进行删除选中的类别。例如,本程序内部对经济类删除采用的是常见词 Business, econom, financ 和 invest。运行程序,当左侧仅填写有 ship, 右侧仅勾选经济类时,处理后的 SCI 期刊数据库工作簿是包含所有含有 ship, 而不含有 Business, econom, financ, invest 的行。

2.4 程序运行及结果

保存处理后的 SCI 期刊数据库工作簿是程序的最后一步。比较弹窗另存和覆盖保存两种方式,后者更为方便。故此,本程序对处理的 SCI 期刊数据库工作簿采取的是覆盖保存方式,因此使用本程序前请确保备份原文件。程序运行成功,不仅会在进度条上有体现,还会以 MsgBox "ok over"方式提醒完成。对于3种处理方式,得到的结果各不相同。

3种方式处理数据对比如表1所示。程序以正选的方式运行成功后,11990条有效待处理数据被筛选保留为488条含有关键词的有效数据。程序运行大约需要2分钟。程序以反删的方式运行成功后,11990条有效待处理数据被筛选删除处理后剩余4932个有效数据,程序运行大约需要7分钟。混合应用程序运行后,11990条有效待处理数据的工作簿被处理生成一个仅剩314行有效数据的工作簿,程序运用大约2分钟。综上可知,正选得到的结果略超出我们需要的内容,反删方式因学科众多不能穷尽关键词的限制其结果大大超出我们的所需,采取混合应用方式对SCI数据库工作簿

进行处理,不仅速度更快,得到的结果也更准确。最后,对混合应用处理后的SCI期刊数据库工作簿进行简单手动处理,即可得到交通运输工程学科所需的182种SCI期刊及其影响因子。原始SCI期刊数据库工作簿及处理后的工作簿可在(<http://pan.baidu.com/s/1o7QkxwI>)获得。

表1 3种处理方式结果比较表

方式	处理所需时间(分)	处理后剩余有效数	有效数据剩余百
		据(行)	分比(%)
正选	2	488	4.07
反删	7	4932	41.13
混合应用	2	314	2.62

3 结语

本程序将 Visual Basic6.0 计算机语言和办公自动化软件 Excel 有机的结合在一起,其混合应用使得科研人员能够方便的得到自己关注的某一两个领域的相关期刊内容及其影响因子,实现了对SCI期刊数据库工作簿的便捷处理,对研究者投稿高水平论文具有一定指导意义。在实际工作中,用VB编写程序对Excel文件进行处理,方法简便易行,且效果颇佳,尤其是对含有大量数据的Excel文件进行处理时,其优点更为显著。

参考文献

- 1 金坤林.如何撰写和发表SCI期刊论文.第2版.北京:科学出版社,2016.
- 2 张显库.Visual Basic 工程应用实例.大连:大连海事大学出版社,2002.
- 3 杨胜,肖崑.Visual Basic 程序设计基础.北京:科学出版社,2016.
- 4 张焰林.Visual Basic 制作 Excel 报表方法探究.计算机系统应用,2008,17(3):111-114.
- 5 王晟.Visual Basic 数据库开发经典案例解析.北京:清华大学出版社,2006.
- 6 Microsoft Visual Basic.MSDN,2008-12.
- 7 张秀香.基于 Android 的健康管理系统客户端的设计与实现.大连:大连理工大学,2012.
- 8 求是科技.Visual Basic6.0 程序设计与开发技术大全.北京:人民邮电出版社,2012.
- 9 方曲祥,吴大中.基于 VB 和 EXCEL 财务报表的简捷实现.计算机应用,1999,19(3):31-33
- 10 姜新.通过 VB 操作 Excel 的应用.无线互联科技,2015,(4):52-54.