

Excel VBA 文献数据处理与分析

宋振世 马宁 牛宁 薛皓 编著



上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

Excel VBA 文献数据处理与分析

宋振世 马宁 牛宁 薛皓 编著



上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

图书在版编目 (CIP) 数据

Excel VBA 文献数据处理与分析 / 宋振世等编著. —上海:
上海科学技术文献出版社, 2018
ISBN 978-7-5439-7657-3

I . ① E… II . ①宋… III . ①表处理软件—应用—文献
数据库—数据库系统 IV . ① TP311.135.9

中国版本图书馆 CIP 数据核字 (2018) 第 127882 号

责任编辑: 王 珺

封面设计: 许 菲

Excel VBA 文献数据处理与分析

Excel VBA WENXIAN SHUJU CHULI YU FENXI

宋振世 马宁 牛宁 薛皓 编著

出版发行: 上海科学技术文献出版社

地 址: 上海市长乐路 746 号

邮政编码: 200040

经 销: 全国新华书店

印 刷: 常熟市人民印刷有限公司

开 本: 787×1092 1/16

印 张: 18.75

字 数: 410 000

版 次: 2018 年 7 月第 1 版 2018 年 7 月第 1 次印刷

书 号: ISBN 978-7-5439-7657-3

定 价: 68.00 元

<http://www.sstlp.com>

内 容 提 要

本书分为文献数据检索、Excel VBA 文献数据处理与分析技术两大部分。旨在帮助读者掌握一定的文献数据检索技术,并进一步对获取的文献数据利用 Excel VBA 技术进行处理,以解决一些人工难以解决的问题。包括了文献检索技术、常用数据库及其检索策略、VBA 基础语法、VBA 运算符与表达式、VBA 程序结构语句、Sub 过程与 Function 函数、VBA 常用函数、Application 对象、Workbook 对象、Worksheet 对象、Range 对象、控制 Word 与 Outlook 程序、VBA 程序界面设计、VBA 数据库技术、处理文本文件、Excel VBA 与 Internet 及文献处理应用案例。

本书把大量知识点应用到现实的工作案例中,使读者掌握知识点的同时,还可以把案例应用到实际工作中,解决实际问题。

本书适合高校图书馆、情报研究等机构开展文献数据处理或学术影响力评价的工作人员;也适合高等学校信息管理类专业学生,作为专业学习的教材使用;也可提供给想掌握 Excel VBA 技术的各类人员,以提高工作效率。

前言

大数据时代,任何决策的制定都是建立在数据分析的基础上,正确、完整的数据可以得出正确的决策,反之会得出错误的决策。显然,数据是最终获得正确决策的基石,在正确、完整的数据基础上才能进一步进行数据分析。

华东师范大学图书馆 2009 年开始面向全校专任教师开展基于文献资源的学术影响力分析,下载 2 000 多位老师的 CSSCI、WOS 核心合集发表论文及论文被引数据,并计算每位老师的 h 指数、发表论文数量、被引频次等几十个评价指标。假设每位老师平均有 10 条文献数据,那么 2 000 多位老师就是两万多条文献记录,虽然数据量不是很大,但是计算每个老师的几十个学术影响力指标,并根据老师的学院、职称进行不同的组合分析,数据量就显得很大。该项工作起初,全部为人工计算各种指标及组合分析,工作人员加班加点利用 Excel 人工处理数据,每天沉浸在 Excel 工作表中,对着电脑“爬格子”,并戏称自己为“表哥”“表姐”。Excel 可视化的操作菜单及其常用功能,几乎可以满足所有的数据需求,只是需要建立在花费大量时间的基础上。学术影响力分析工作虽然劳累,但最终形成的分析报告得到了相关领导的重视,随后每年继续开展此项工作,分析的数据逐渐扩展到教学、获奖、培养学生和项目等统计内容。随着数据量的激增,仅靠人工操作 Excel 很难准确、高效地完成工作。2011 年,我们开始探索利用 Excel VBA 技术实现文献数据处理与分析工作,从文献数据的获取、处理、分析和形成报表来看,Excel VBA 技术帮助实现了很多人工难以完成的工作,如为 SCIE 论文抓取中科院期刊分区、影响因子等数据,原来需要几天才能完成的工作,通过 Excel VBA 技术编写的宏程序仅需要几分钟的时间,而且数据的准确性较高。

2016 年科睿唯安“双一流建设中的数据分析与应用论坛”中,约 250 多位来自逾百所大学的图书馆和科研管理部门人员参会,多家高校图书馆和科研管理部门介绍科研评价或评估工作。可见,高校图书馆和科研管理部门已纷纷开展基于文献资源的决策支持服务,然而开展该项服务的工作人员大多为非专业的 Excel 编程人员,但大多都具备一定的 Excel 的操作技能和简单的编程基础。基于此,作者编写了该书,以帮助高校图书馆和科研管理部门工作人员更好地掌握 Excel VBA 技术,提高工作效率。

该书分为文献数据检索、Excel VBA 文献数据处理与分析技术两大部分。旨在帮助

读者掌握一定的文献数据检索技术,并进一步对获取的文献数据利用 Excel VBA 技术进行处理,以解决一些人工难以解决的问题,该书适合高校图书馆、情报研究等机构开展文献数据处理或学术影响力评价的工作者,也适合高等学校信息管理类专业学生,作为专业学习的教材使用。该书包括了文献检索技术、常用数据库及其检索策略、VBA 基础语法、VBA 运算符与表达式、VBA 程序结构语句、Sub 过程与 Function 函数、VBA 常用函数、Application 对象、Workbook 对象、Worksheet 对象、Range 对象、控制 Word 与 Outlook 程序、VBA 程序界面设计、VBA 数据库技术、处理文本文件、Excel VBA 与 Internet。

该书的主要特色:一、深入浅出地介绍了 Excel VBA 技术并穿插了文献数据处理部分案例;二、在总结华东师范大学图书馆开展的学术影响力评价工作实践的基础上,形成了大量可直接使用的案例,读者可以直接使用这些案例中的宏程序开展工作,以提高工作效率。

华东师范大学图书馆情报咨询部所有同事在利用该书中编写的宏程序过程中提出了宝贵的改进建议,同时该书受华东师范大学学术评价与促进研究中心和华东师范大学教师绩效数据库项目资助,在此一并表示感谢!

书中所有案例已在 Excel2013 版本中测试利用。

由于作者水平有限,书中缺点、错误在所难免,敬请读者批评指正,不吝赐教。

作 者

2017 年 12 月

目 录

第一部分 文献数据检索

第一章 文献检索技术	3
1.1 布尔逻辑检索技术	3
1.2 位置限定运算符 SAME	3
1.3 截词检索技术	4
1.4 高级检索	4
第二章 常用数据库及其检索策略	7
2.1 中文社会科学引文索引(CSSCI)	7
2.2 中国科学引文数据库(CSCD)	8
2.3 Web of Science 核心合集	9
2.4 基本科学指标数据库(ESI)	11
2.5 JCR 期刊分区表	12
2.6 InCites 数据库	13

第二部分 Excel VBA 文献数据处理与分析技术

第一章 打开 VBA 的大门	17
1.1 关于 VBA	17
1.2 VBA 在 Excel 中的优势	17
1.2.1 批量执行任务	17
1.2.2 完成 Excel 本身无法完成的任务	17
1.2.3 开发专业程序	17
1.3 VBA 和宏	18
1.3.1 录制和保存宏	18
1.3.2 编辑修改录制的宏	20
1.3.3 执行宏	23
1.4 宏的安全性	24

1.4.1	打开包含宏的文档	24
1.4.2	设置宏的安全性	24
1.5	VBA 开发环境	26
1.5.1	VBE 的启动方式	26
1.5.2	VBE 操作界面	27
1.5.3	编写和调试程序	30
1.6	VBA 编程思想	33
1.6.1	面向对象的编程	33
1.6.2	Excel 中的对象	35
第二章	VBA 基础语法	36
2.1	VBA 标识符	36
2.2	VBA 关键字	37
2.3	VBA 常量	37
2.4	VBA 变量	37
2.5	VBA 基本数据类型	39
第三章	VBA 运算符与表达式	42
3.1	算术运算符	42
3.2	比较运算符	42
3.3	连接运算符	43
3.4	逻辑运算符	44
3.5	运算符的优先级	44
第四章	VBA 程序结构语句	46
4.1	顺序结构语句	46
4.1.1	赋值语句	46
4.1.2	输入语句	48
4.1.3	输出语句	50
4.2	分支结构语句	53
4.2.1	if ... then	53
4.2.2	if ... then ... else	53
4.2.3	if ... then ... elseif	54
4.3	循环结构语句	55
4.3.1	For ... Next	55
4.3.2	嵌套 For ... Next	56
4.4	错误处理语句	58
4.4.1	On Error 语句	59
4.4.2	Resume 语句	60

第五章	Sub 过程与 Function 函数	61
5.1	过程的分类	61
5.1.1	事件过程	61
5.1.2	Sub 过程	63
5.1.3	Function 函数过程	64
5.2	调用过程和函数	65
5.2.1	调用 Sub 过程	65
5.2.2	调用 Function 函数过程	65
5.3	常用文献分析自定义函数	67
5.3.1	统计论文作者数	67
5.3.2	案例: ht 指数	68
5.3.3	字体过程函数	70
5.3.4	创建表格过程函数	71
第六章	VBA 常用函数	72
6.1	字符串处理函数	72
6.1.1	Len 函数	72
6.1.2	Left、Right 和 Mid 函数	72
6.1.3	Trim、LTrim 和 RTrim 函数	72
6.1.4	LCase 和 UCase 函数	73
6.1.5	Space 函数	73
6.2	日期函数	74
6.2.1	Time、Now 和 Date 函数	74
6.2.2	Year、Month、Day、Hour、Minute 和 Second 函数	74
6.3	格式化函数	75
6.3.1	Format 函数	75
6.3.2	FromatDateTime 函数	79
第七章	操作 Application 对象	80
7.1	Application 对象概述	80
7.2	Application 对象属性	80
7.3	返回对象	82
7.4	常用方法	84
7.5	Application 对象事件	85
7.6	工作表函数(WorksheetFunction)	86
第八章	操作 Workbook 对象	87
8.1	引用 Workbook 工作簿对象	87
8.1.1	通过索引号引用工作簿	87

8.1.2 通过名称引用工作簿 87

8.1.3 引用活动工作簿 87

12.2	自定义功能区的方法	120
12.3	RibbonX 自定义功能区	121
12.4	CommandBars 对象自定义功能区	126
12.5	自定义工具栏	129
12.6	自定义全新快捷菜单	132
12.7	自定义内置快捷菜单	138
12.8	自定义快速访问工具栏	140
第十三章	VBA 数据库技术	145
13.1	ADO 数据库访问技术	145
13.1.1	Connection 对象	146
13.1.2	Recordset 对象	147
13.1.3	使用 ADO 访问数据库的步骤	150
13.2	SQL 语言	152
13.2.1	Select 查询语句	152
13.2.2	Update 修改语句	153
13.2.3	Delete 删除语句	153
13.2.4	Insert 插入语句	153
13.2.5	利用 Like 运算符进行模糊查询	154
13.3	SQL 函数	154
13.4	利用 ADO 和 SQL 操作 Excel 工作簿	155
13.4.1	查询当前工作簿中其他工作表的数据	156
13.4.2	查询其他指定工作簿中的数据	157
第十四章	处理文本文件	158
14.1	常用文件操作语句	158
14.1.1	Dir 函数查找文件	158
14.1.2	Open 语句打开文件	159
14.1.3	Name 语句重命名文件	159
14.1.4	Kill 语句删除文件	160
14.1.5	FreeFile 函数获取文件号	161
14.1.6	Close 语句关闭文件	161
14.1.7	Input # 语句	161
14.1.8	Eof 函数	162
14.1.9	Line Input 语句读出文件数据	162
14.1.10	FileCopy 复制文件语句	162
14.2	案例：汇总 CSSCI 来源文献	163

第十五章 Excel VBA 与 Internet	167
15.1 用 VBA 创建超链接	167
15.2 案例: CNKI-E Study 论文添加超链接	168
15.3 Web 查询获取 Internet 上的数据	169
15.4 了解 XmlHttp 对象和 QueryTable 对象	172
15.4.1 XmlHttp 对象	172
15.4.2 QueryTable 对象	175
15.5 案例: Web 查询获取 CNKI 引证文献	176
第十六章 文献处理与分析应用案例	178
16.1 案例: 通用 Excel 数据查询汇总	178
16.2 案例: 通用 Excel 数据复制粘贴汇总	180
16.3 案例: SCIE 论文 txt 格式文件数据汇总	183
16.4 案例: CSSCI 被引文献汇总	186
16.5 案例: SCIE 论文标题大小写转换	192
16.6 案例: 万方专利数据格式转换	194
16.7 案例: 获取中科院分区表数据	197
16.8 案例: 构建影响因子数据库	202
16.9 案例: 自定义分组统计	204
16.10 案例: SQL 聚合函数统计	208
16.11 案例: Word 分析报表	210
16.12 案例: h 指数的统计	218
16.13 案例: SCIE 论文查收查引系统	223
16.14 案例: 中文作者 SCIE 论文 author 检索式	234
16.15 案例: 通过姓名匹配 SCIE 论文	238
16.16 案例: 删除字符串内多余的空格	244
16.17 案例: CSSCI 被引数据合并	244
16.18 案例: 显示程序运行进度	248
16.19 案例: 提取论文作者机构	252
16.20 案例: 科研方阵划分	253
16.21 案例: 百度一下	254
16.22 案例: 共词矩阵	255
16.23 案例: ESI 预测	258
16.24 案例: 论文数据查重	262
16.25 案例: 任意分组统计	263
16.26 案例: 获取中国知网论文被引数据	267
16.27 案例: 数据对比	272

16.28 案例：职称论文数据报表 273

16.29 案例：SQL 语句查询中文字排序..... 275

16.30 案例：灵活运用程序代码 276

参考文献 277

附录 A 表目录 278

附录 B 图目录 280

第一部分

文献数据检索

第一章 文献检索技术

1.1 布尔逻辑检索技术

布尔逻辑运算符包括 AND(与)、OR(或)和 NOT(非),用来表示两个检索项之间的逻辑关系,AND 表示其所连接的两个检索项的交叉部分,即存在交集的部分;OR 用于连接并列关系的两个检索项,即存在和集的部分;NOT 用于连接排除关系的检索项,即排除不需要的和影响检索结果的检索项。A 和 B 分别代表两个检索词,三种布尔逻辑运算符的逻辑关系如下表:

表 1 布尔逻辑运算符逻辑关系图

逻辑运算符	AND(与)	OR(或)	NOT(非)
检索式	A AND B	A OR B	A NOT B
检索结果	A 和 B 检索词都出现的文献记录	A 和 B 检索词任意出现的文献记录	只出现 A,而不出现 B 文献记录

在数据库中,常用布尔运算符连接多个检索词检索,AND 可以缩小检索范围,有利于提高查准率;OR 可以扩大检索范围,防止漏检,有利于提高查全率。

1.2 位置限定运算符 SAME

在“地址”检索中,使用 SAME 将检索限制为出现在“全记录”同一地址中的检索词。需要使用括号来分组地址检索词,例如:(McGill Univ SAME Quebec SAME Canada)查找在“全记录”的“地址”字段中出现 McGill University 以及 Quebec 和 Canada 的记录。当在其他字段(如“主题”和“标题”)中使用时,如果检索词出现在同一记录中,SAME 与 AND 的作用就完全相同。例如:(cat SAME mouse)与(cat AND mouse)将得到相同的结果。

使用 SAME 运算符的原因是,它指定了由运算符连接的检索词位于同一地址,而不仅仅是位于相同的字段。使用 SAME 运算符限制您的检索。例如,检索式 IBM SAME NY 检索全记录“地址”字段(或“通讯作者地址”字段)中包含这两个检索词的记录。例如:IBM Res Corp, Yorktown Heights, NY 10598 USA。

在“地址”检字段索中,使用 AND 运算符可将检索范围扩大到记录中的所有地址,一条记录中包含多条地址。检索式 IBM AND NY 可检索到 IBM SAME NY 可检索的所有记录。此外,它检索 IBM 和 NY 位于“地址”字段中,但属于不同地址的记录。例如,IBM AND NY 将检索包含这两个地址的记录: 1. Cornell Univ, Dept Mat Sci & Engr, Ithaca, NY 14853 USA; 2. IBM Corp, Almaden Res Ctr, Div Res, San Jose, CA 95120 USA。可以看出,返回 NY 在一个地址字段出现,且 IBM 在不同地址字段出现的记录。

1.3 截词检索技术

使用通配符来进行截词检索,通配符表示未知字符,仅在英文查询中有效。通配符星号(*)表示任何字符组,包括空字符;问号(?)表示任意一个字符;美元符号(\$)表示零或一个字符。姓名中的连字符(-)和撇号(')视为空格处理。不同数据库中,通配符的使用规则略有不同。以 Web of Science 核心合集为例,其通配符的使用规则如下:

- 在 Web of Science 的所有产品数据库中,可以在如下检索字段中使用左截词符:“主题”“标题”“入藏号”和“识别代码”。
- 在“主题”和“标题”检索中,如果使用左截词符,那么必须在通配符后至少输入 3 个字符。例如: * bio。
- 在“入藏号”和“识别代码”检索中,如果使用左截词符,那么必须在通配符后至少输入 1 个字符。例如: * 2307。
- 在“主题”和“标题”检索中,如果使用右截词符,那么必须在通配符前至少输入 3 个字符。例如: bio *。
- 在“作者”检索中,检索姓氏时必须在通配符前输入至少 2 个字符,例如: sm *
- 通配符可位于检索词的中间。例如,odo\$r可查找到 odor 和 odour。
- 您不能在以下字符后面使用通配符: 特殊字符 (/ @ #)和标点符号(., : ; !)。
- 您不能在出版年检索中使用通配符。例如,可以使用 2007,但不能使用 200 *。
- 如果通配符出现在单词或名称中,您就不能对其进行检索。例如,检索式 TS=E * Trade OR TS=“E * Trade”无法返回有关此公司的记录。
- 不能在引起的检索内使用美元符号(\$)。
- 在截词后匹配范围很广的情况下,请尽量避免使用通配符。例如,检索式 UT=* 2或 UT=* 2 * 或 UT=* 22 或 UT=* 22 * 返回的结果可能并不完全(或者没有结果),这是因为产品数据库中与之匹配的记录太多。

1.4 高级检索

高级检索是利用布尔逻辑运算符连接多个检索字段,实现多个字段组合检索。以 Web of Science 核心合集平台为例,可以利用高级检索功能检索某一作者在某个机构的