七年级（上）信息技术第13课时

编辑数据（一）学程拓展

在整理数据的时候，同学们会发现拼写错误的单词、难以去除的尾随空格、不需要的前缀、不正确的大小写和非打印字符给人一种不好的第一印象。导致数据混乱的因素还不止这些。请拓展阅读资料，了解更多清理数据的方法吧。

**清理数据的基础知识**

你并不一定始终可控制从数据库、文本文件或网页等外部数据源导入的数据格式和类型。通常需要先清理数据，才能分析数据。幸运的是，Excel 提供许多功能，可帮助你获取所需精确格式的数据。有时任务非常简单，Excel 具有执行此任务的特定功能。例如，可轻松使用拼写检查清理包含批注或说明的列中拼写错误的单词。或者如果想要删除重复行，可使用“**删除重复项**”对话框快速执行此操作。

在其他情况下，可能需要使用公式将导入的值转换为新值来操作一列或多列。例如，如果想要删除尾随空格，可创建新列来清理数据，方法是：使用公式，向下填充新列，将新列的公式转换为值，然后删除原始列。

清理数据的基本步骤如下所示：

1. 从外部数据源导入数据。
2. 在单独的工作簿中创建原始数据的备份副本。
3. 确保数据采用行和列的表格格式：每一列中的数据类似、所有列和行可见并且该区域中无空白行。为获得最佳效果，请使用 Excel 表格。
4. 先执行不需要列操作的任务，例如拼写检查或使用“**查找和替换**”对话框。
5. 接下来，执行需要列操作的任务。列操作的常规步骤有：
	1. 在需要清理的原始列 (A) 旁插入新列 (B)。
	2. 在新列 (B) 的顶部添加将转换数据的公式。
	3. 在新列 (B) 中向下填充公式。在 Excel 表中，会使用向下填充的值自动创建计算列。
	4. 选择新列 (B)，将其复制，然后作为值粘贴到新列 (B) 中。
	5. 删除原始列 (A)，这会将新列从 B 转换为 A。

若要定期清理相同的数据源，请考虑录制宏或编写代码，自动执行整个过程。

**删除重复行**

导入数据时，重复行是一个常见问题。最好先筛选唯一值，确认结果是所需结果，然后再删除重复值。

**查找和替换文本**

可能需要删除常见的前导字符串（例如后跟冒号和空格的标签）或后缀（例如已过时或不必要的字符串结尾处的附加说明短语）。若要执行此操作，可查找文本的实例，然后将其替换为无文本或其他文本。

**更改文本大小写**

有时文本格式混乱，尤其是文本大小写方面。使用三种 Case 函数中的一种或多种，可将文本转换为小写字母（如电子邮件地址）、大写字母（如产品代码）或首字母大写（如姓名或书名）。

**删除文本中的空格和非打印字符**

有时文本值包含前导空格、尾随空格或多个嵌入空格字符（Unicode 字符集值 32 和 160），或非打印字符（Unicode 字符集值 0 到 31、127、129、141、143、144 和 157）。执行排序、筛选或搜索操作时，这些字符有时会导致意外结果。例如，在外部数据源中，可能会无意添加额外的空格字符，从而导致打字错误，或者从外部源导入的文本数据可能包含嵌入在文本中的非打印字符。由于这些字符不容易引起注意，因此意外结果可能很难理解。若要删除这些不需要的字符，可组合使用 TRIM、CLEAN 和 SUBSTITUTE 函数。

**修复数字和数字符号**

主要有两个数字问题可能需要你进行数据清理：无意中将数字导入为文本，以及需要根据你组织的标准更改负号。

**修复日期和时间**

由于存在许多不同的日期格式，并且这些格式可能混杂有编号部件代码或其他包含斜杠标记或连字符的字符串，因此日期和时间通常需要进行转换和重新设置格式。

**合并和拆分列**

从外部数据源导入数据后的常见任务是将两列或多列合并为一列，或将一列拆分为两列或多列。例如，可能需要将包含全名的列拆分为名字和姓氏。或者可能需要将包含地址字段的列拆分为单独的街道、城市、地区和邮政编码列。反之亦可。可能需要将名字和姓氏列合并为一个全名列，或者将单独的地址列合并为一列。其他可能需要合并为一列或拆分为多列的常见值包括产品代码、文件路径和 Internet 协议 (IP) 地址。

**转换和重新排列行和列**

Office Excel 中的大多数分析和格式设置功能都假设数据存在于单个平面二维表中。有时可能需要将行转换为列、将列转换为行。有时候，数据甚至不是表格格式结构，需要使用一种方法将数据从非表格格式转换为表格格式。

**通过联接或匹配协调表格数据**

有时，数据库管理员会使用 Office Excel 查找并更正两个或多个表联接时的匹配错误。这可能涉及协调不同工作表中的两个表，例如，查看两个表中的所有记录，或比较两个表并查找不匹配的行。