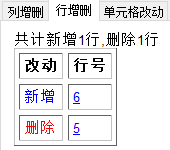
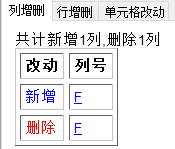
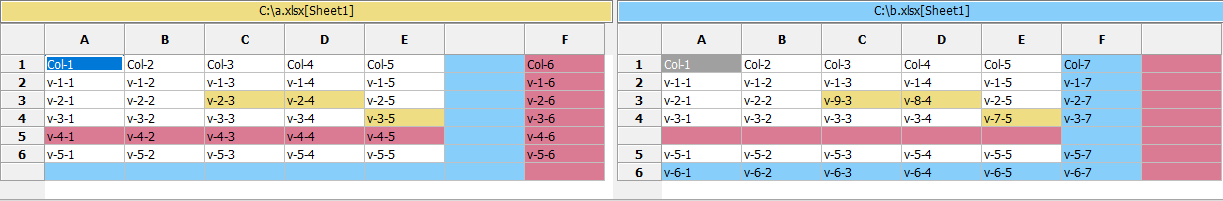
##### 软件需求说明书（软件规格说明书，系统测试需要的标准文档）：

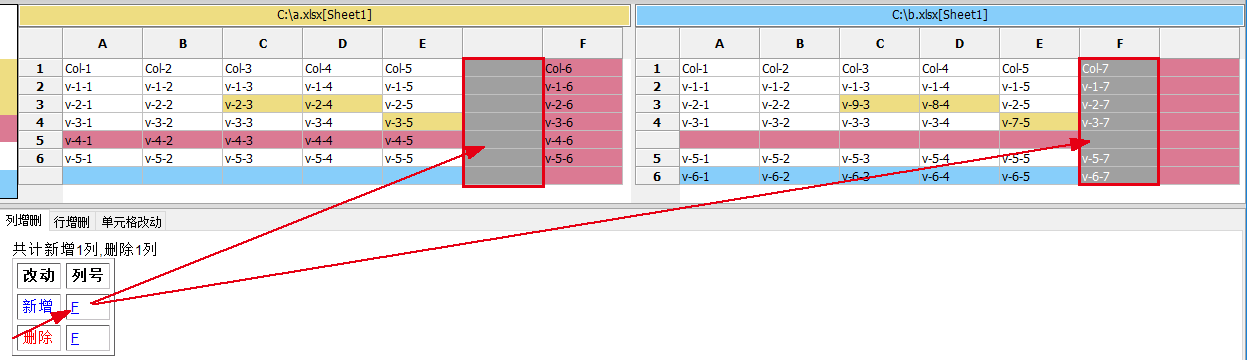
1. 请提供交互界面让用户指定两个用于比较的Excel文件；
2. 依据sheet名来计算和展示两个Excel文件中sheet表的增删改；自动比对两个Excel文件中任意两个sheet名一致的sheet表内容差异；
3. 请设计并实现一个实用算法，用于比对两个Sheet之间的差异，需要甄别计算出以下差异：**列增删**、**行增删**、**单元格改动**；



1. 计算出来的不同差别内容能够在界面上面用不同颜色标识展示，方便用户直接识别（如下图，红色代表删除，浅蓝色代表新增，黄色代表修改；另外的灰色和深蓝色代表被选定）；



1. 差别内容和展示面板需要实现联动，列出的差异内容提供点击导航功能（如下图，点击下方差异点后能在表格比对面板中马上定位）



##### 概要设计说明书：

本软件基于Python 2.7，使用PYQT4构建GUI，使用PYINSTALLER打包为Windows平台下可执行文件（.exe）。

此软件，主要包含三个文件：

GUI.py:软件界面的实现，主要包含部件的创建，布局管理等。

界面主要分成三部分，最上面部分是选择文件按钮和开始diff动作按钮，其diff结果展示在下面部分。下面部分分成左右两侧，左侧展示文件包含的表名，点击表名会在右侧展示具体结果。

规定，新增的行或列用蓝色表示，删除的行或列用橘色表示，改动的单元格用黄色表示，未改动的行或列或单元格用白色表示。

ExcelDiff.py:主要实现两个Excel文件的diff功能。

输入：两个Excel文件的地址；

输出：包含diff结果的list，其中每一项有三部分组成：表名，

展示表1的数据，展示表2的数据、改动信息（改动的行号，改动的列号，改动的单元格号）。

逻辑流程主要是，将excel文件按照行或列的形式转化为list，用diff.py（后面详细说明）比较两个list，将结果进行适当匹配，寻找相似性高的行（列）设为一行通过修改后成为另一行。最后对数据进行适当的组织以适用界面需求。

diff.py:实现两个list的diff。

输入：两个list，以及是否是表格数据

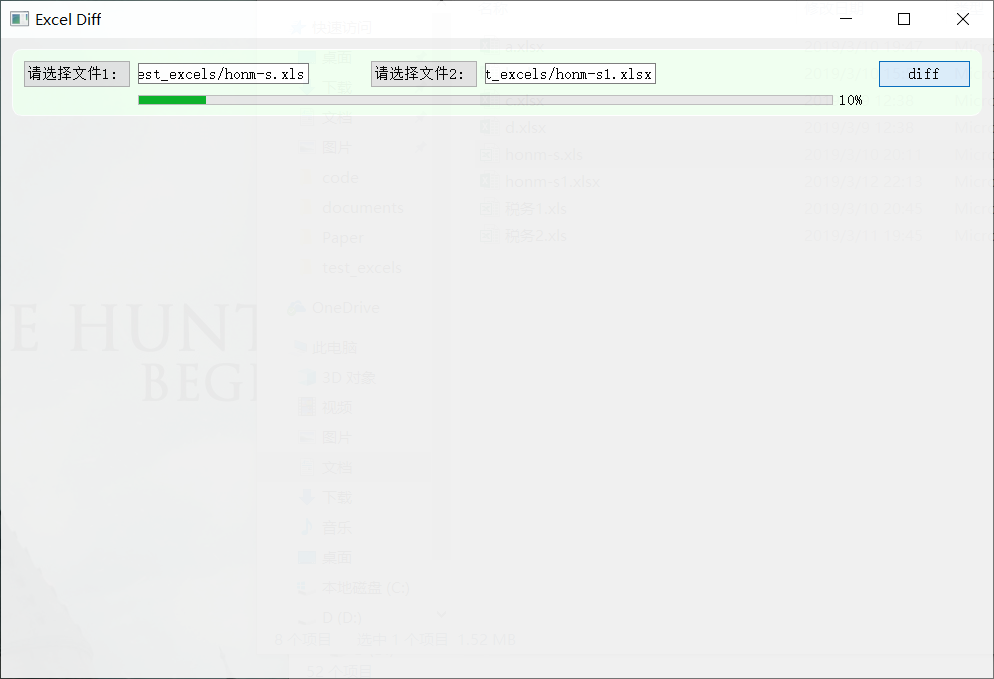
输出：标记list的每个元素为新增或者删除或者改动或者未修改。

基于在大多数情况下，大部分数据相同的假设，也就是两个Excel是通过修改其中一个获取另一个而来，那么对diff算法做了部分优化，例如提取公共前缀、公共后缀，包含关系，两端不同中间相同时分成两部分各自比较等方法。最后，使用LCS算法计算两个list的变化过程，获取diff结果。

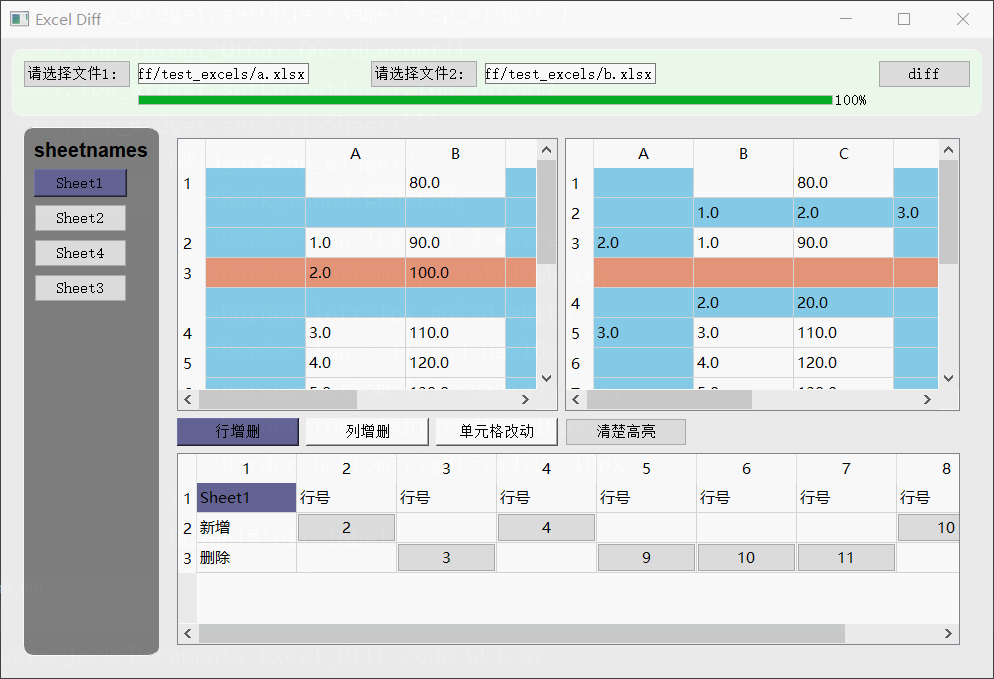
规定两行（列）之间的未改动单元格的比例少于50%时视为不相干两行（列），低于50%视为两行相似。

##### 用户操作手册：

　　点击选择文件按钮选择文件，然后点击diff按钮开始diff。



等待进度条读完后，在下方显示详细结果。点击相应表名，可查看具体结果。



##### 测试：

在生产过程中，对于每项功能，单独进行测试没有发现问题。后进行联合测试亦没有发现问题。

对于大型文件的处理能力，本人用包含两个sheet，每个sheet有2000行\*20列的Excel进行测试，能在1分钟内给出结果。

##### 项目开发总结：

项目开发过程中，进行了多次优化，其中结合实际生产过程的优化最为明显。合理安排事件处理的顺序、高发生概率的事件做优先处理等可以明显提升软件处理的速度，因此可以说一份清晰、明确的需求是多么重要。

同时，项目开始时没有做好充分设计便开始编写，导致后期推倒重做多次，浪费了大量时间，是一段惨痛的教训。